Schlußbericht

Verbundprojekt

Multimediale wissens- und fallbasierte Trainings- und Informationssysteme in der Medizin

Gefördert vom Bundesminister für Bildung und Forschung beim
Projektträger
Deutsches Zentrum für Luft- und Raumfahrt (DLR)
Förderkennzeichen: 01EI9603/1

Dr. Bettina Reinhardt
Prof. Dr. Frank Puppe
Lehrstuhl für Künstliche Intelligenz und Angewandte Informatik
der Universität Würzburg
Am Hubland
97074 Würzburg

Prof. Dr. Kurt Kochsiek (erem.)
Medizinische Klinik
der Universität Würzburg
Josef-Schneider-Straße
97080 Würzburg

Prof. Dr. Stefan Schewe
Medizinische Poliklinik
der Ludwig-Maximilians-Universität München
Pettenkoferstr. 8a
80336 München

Inhaltsverzeichnis

1

Zusammenfassung

Ziel des Vorhabens war die Ausarbeitung einer Methodik zur Entwicklung multimedialer, wissensbasierter Trainings- und Informationssysteme, deren Umsetzung in einem wiederverwendbaren Diagnostik-Expertensystem-Shell-Baukasten, der Aufbau von Trainingssystemen in verschiedenen medizinischen Teilbereichen und deren Evaluation und Publikation in einem Verlag. Der Endbenutzer der Trainingssysteme soll aus vielen multimedial präsentierten Kasuistiken (mehrere pro Diagnose) auswählen können, bei denen er die Symptome erkennen und interpretieren sowie Tests zur Abklärung der Diagnose auswählen mufl. Diese Aktionen sowie ihre Begründungen werden dabei vom Trainingssystem durch Vergleich mit dessen Vorgehensweise kritisiert. Ausgangspunkt des Vorhabens ist der funktionsfähige, erfolgreich eingesetzte Shell-Baukasten D3 zur Entwicklung nicht-multimedialer, wissensbasierter Diagnose, Informations- und Trainingssysteme in der Medizin.

Der Shell-Baukasten D3 wurde um die Komponente des D3-Trainers erweitert, die das problembasierte Lernen unterstützt und verschiedene didaktische Gestaltungselemente bezüglich Interaktivität, Präsentation und Kritik der Aktionen des Lernenden anbietet. Damit kann der Fachexperte nicht nur den Inhalt des Trainingssystem eingeben, sondern auch seine didaktischen Ziele umsetzen. Die wichtigsten Aktionsmöglichkeiten, die der D3-Trainer anbietet, sind:

- Auswahl von Verdachtsdiagnosen und Therapien,
- Begründen von Verdachtsdiagnosen und Therapien,
- Auswahl zusätzlicher Tests / Untersuchungen,
- Begründung der Auswahl von Tests / Untersuchungen,
- Erkennen von Symptomen auf Bildern,

- Hervorheben der Bildregionen, auf denen die Symptome erkannt wurden.

Das nächste Kapitel enthält nach einer kurzen Motivation für computerbasierte Trainingssysteme eine ausführliche Übersicht über fallbasiertes Lernen in der Diagnostik und deren Umsetzung im D3-Trainer. Im dritten Kapitel wird der D3-Trainer sowohl aus Autoren- als auch aus Anwendersicht detailliert beschrieben. Anschlieflend werden der RheumaTutor (erschienen als CD: S. Schewe: RheumaTutor - fallbasiertes rheumatologisches Trainingsprogramm, Ullstein Medical, 1999, ISBN: 3-86126-934-1) und andere mit dem D3-Trainer entwickelte Trainingssysteme vorgestellt und die Ergebnisse der durchgeführten Evaluationen präsentiert. Ein Ausblick auf sinnvolle Weiterentwicklungen des D3-Trainers folgt im fünften Kapitel. Im Anhang werden Originaldaten zur Evaluation der Rheumatrainers präsentiert.

2

Motivation und Übersicht

2.1 Computerunterstütztes Lernen

„Das Wissen der Menschheit verdoppelt sich alle 5 Jahre, während sich die Halbwertszeit des Wissens ständig verringert“[SERVE 1992]. Eine Folge davon ist, daß man in der heutigen Zeit nicht mehr davon ausgehen kann, mit einer vorhandenen Ausbildung sein Leben lang bestehen zu können. Vielmehr werden alle Bereiche des Alltags von einem *lebenslangen Lernen* bestimmt sein. Eine mögliche Unterstützung bei der Lösung dieses Problems können *Computer–Basierte Trainingssysteme* (CBT) sein, die das Medium Computer seinen Vorteilen und Fähigkeiten entsprechend einsetzen.

Der Computer als Lehrer bietet einige positive Voraussetzungen, um in einem globalen Lernszenario eine wichtige Rolle spielen zu können. Wichtig für die Lernsituation ist vor allem, in welcher sozialen Lernumgebung der Student sich befindet. Grundsätzlich können dabei die folgenden Szenarien unterschieden werden: An dem *Einzel–Lernplatz* ist der Lernende in seiner Lernumgebung isoliert und kann aber auch gerade deshalb sehr individuell auf die eigenen Bedürfnisse ausgerichtet lernen. Das entsprechende herkömmliche Medium ist hier das konventionelle Lehrbuch oder Nachschlagewerk, mit dem sich der Lernende neues Wissen aneignen kann. Der Einzel–Lernplatz ist aber auch besonders zur Nacharbeit und Wiederholung von Wissen geeignet, welches in anderen Lernumgebungen erworben wurde. Hier findet der Computer eine gute Anwendung, da er sich auf individuelle Fehler, Lerntempo oder Ähnliches des Lernenden einstellen und diesen besser unterstützen kann als ein Buch es könnte.

In der *Partnerarbeit* oder in der Arbeit in *Kleingruppen* kann es zu sehr guten Ergebnissen kommen, wenn die Partner sich gut verstehen. In solch kleinen Gruppen können aber auch persönliche Aversionen einen Lernerfolg fast vollständig ausschließen. Das Lernen in dieser Umgebung erfolgt

meist ungeführt, d. h. ohne einen Tutor. Die in dieser Umgebung entstehenden Diskussionen können zu einem tieferen Verständnis des behandelten Stoffes führen und stärken außerdem das soziale Verhalten der Beteiligten. In der Partnerarbeit kann der Computer nur beschränkt eingesetzt werden, so beispielsweise als Diskussionsgrundlage bei einer Problembearbeitung.

Sind die Lernenden in einer großen Gruppe einem Lehrenden gegenübergestellt, wie z. B. in einer Schulstunde oder einer Vorlesung, wird dies als *Frontalunterricht* bezeichnet. Die Interaktivität zwischen Lehrenden und Lernenden ist auf ein Minimum beschränkt, der Lehrinhalt ist Wissensvermittlung mit rein darbietenden Techniken. Hier kann der Computer als ergänzendes Element in den Unterricht eingebaut werden, beispielsweise für Präsentationen vor der Gruppe.

Geht man davon aus, daß der Lernerfolg am größten ist, wenn genau ein Lehrer für genau einen Schüler zuständig ist, kann ein didaktisch sinnvoll gestaltetes Lernprogramm diese 1:1–Lernumgebung nachbilden. Durch die Beschränkung auf nur einen Lernenden kann dieser selbst ein angemessenes Lerntempo wählen. Auch sind die Lernschwerpunkte von dem Lernenden selbst abhängig, d. h. er verweilt länger bei einem Thema, welches ihn mehr interessiert oder mehr fordert, da es für ihn persönlich schwieriger ist. Durch die computerbasierte Lehre ist es auch für jeden Lernenden möglich, an dem Ort und zu der Zeit zu lernen, die für ihn optimal ist, ohne durch feste Raum– und Lehrpläne beschränkt zu werden.

Ein computerbasiertes Lernprogramm kann eine dynamische Diskussionsgrundlage bieten und das gemeinsame Problemlösen kooperative und kollaborative Prozesse anstoßen. Der Computer als Grundlage für ein diskussionsfreudiges Gruppenlernen unterstützt ein mehr sozial orientiertes Lernen, welches in der momentanen Gesellschaft gefordert wird. Nicht nur Fachwissen, sondern auch soziale Kompetenz können so in der Gruppe geschult werden.

Ein Lehrer, der in seinen Unterricht computerbasierte Lernprogramme einbaut, erreicht einen höheren Attraktivitätsgrad durch die Hinzunahme neuer Medien. Gleichzeitig kann auch eine Entlastung des Lehrers stattfinden, der sich auf Schwerpunkte konzentrieren kann, die entweder gar nicht oder nur schwer durch den Computer vermittelt werden können.

Ein weiterer Vorteil des CBT ist die Tatsache, daß beim Üben einer Tätigkeit am Computer kein Personen– oder Sachschaden entstehen kann. So können z. B. gefährliche Tätigkeiten am Computer eingeübt werden, ohne die Lernenden oder andere zu gefährden. Beispiele für solche Tätigkeiten sind die Durchführung eines chemisches Experiment mit ätzenden Stoffen oder die Diagnostik und Therapie eines Patienten in der Medizin. Ein weiterer Einsatz solcher Systeme findet sich in Anwendungen, bei denen eine physikalische Simulationsumgebung zu teuer ist und nur zum En-

de eines Ausbildungskurses angeboten wird, wie bei Simulationsanlagen technischer Systeme (z. B. Kraftwerke, Flugzeuge).

Zusammenfassend kann man also folgende Gründe für den Einsatz von Computern in der Aus– und Weiterbildung nennen:

- Unabhängiger Lernzeitpunkt
- Unabhängiger Lernort
- Eigenes Lerntempo
- Möglichkeit, eigene Schwerpunkte zu setzen
- Individuelle Lehrstrategien
- Vielfältige Lernmethoden
- Unterstützung von sozialen Aspekten
- Entlastung der Lehrkraft
- Lernen gefährlicher Tätigkeiten am realistischen Modell
- Alternative zu teuren Ausbildungssimulatoren

Ein didaktisches Modell, welches in den letzten Jahren immer häufiger zum Einsatz kommt, ist das problemorientierte Lernen. In diesem Ansatz wird eine Problemstellung zum Ausgangspunkt des Lernprozesses. Das Problem wird in Kleingruppen und im Selbststudium diskutiert und eine Lösung dafür erarbeitet. Diese Art der Lehre kann besonders gut durch den Computer unterstützt werden.

2.2 Fallbasierte Trainingssysteme in der Diagnostik

Für den Einsatz im problemorientierten Lernen besonders gut geeignet sind fallbasierte Trainingssysteme, die es dem Lernenden ermöglichen, aktiv an einer Problemlösung zu arbeiten. Dazu wird ihm ein Fall präsentiert, an dem er idealerweise alle Aktionen ausführen kann, die er auch in der realen Problemlöseumgebung zur Verfügung hat.

Problemorientiertes Lernen kann in vielen Anwendungsszenarien auftreten, in dieser Arbeit wird aber nur der Bereich der diagnostischen Problemlösung betrachtet. Bei der Beschränkung der Problemszenarien auf diagnostische Domänen, wie beispielsweise medizinische Diagnostik oder Fehlererkennung in Maschinen, stellen sich folgende Aufgaben für den Diagnostiker:

- Merkmalserhebung,
- Merkmalserkennung,
- Merkmalsinterpretation,
- Problembehebung.

Bei der Merkmalserhebung geht es um die Auswahl neuer Untersuchungen oder Tests, die für die weitere Problemlösung benötigt werden. Manche Merkmale können nicht direkt abgelesen werden, sondern müssen aus einem Bild, Video oder Ton erkannt werden. Beispiel für diese Merkmalserkennung ist das Feststellen eines Herzfehlers aus dem Abhören des Patienten oder die Festlegung auf einen Druckfehler wie Dublieren durch Ansehen des Produktes. Das Schlußfolgern auf Lösungen basierend auf diesen Merkmalen wird hier mit Merkmalsinterpretation bezeichnet. Bei der Problembehebung wird aufgrund der gefundenen Lösung eine Aktion gewählt, die das Problem des Falles behebt.

Diese Aufgaben werden von intelligenten Tutorsystemen gelöst, indem das zugrundeliegende Wissen in expliziter (deklarativer) Form repräsentiert und aus diesem Wissen die Lehrumgebung generiert wird. Intelligente Tutorsysteme werden typischerweise in vier Komponenten (Benutzerschnittstelle, Studentenmodell, Didaktik und Wissensmodell; siehe auch Abb. **??**) aufgeteilt. Tabelle 2.1 gibt eine Übersicht über wichtige Aspekte dieser Komponenten in Bezug auf fallbasierte Trainingssysteme in der Diagnostik. Ein innovatives System, welches in dieser Arbeit ausgearbeitet wurde, ist der D3TRAINER, mit dem intelligente fallbasierte Trainingssysteme generiert werden können. Wie der D3TRAINER die aufgelisteten Aspekte behandelt, wird ebenfalls in der Tabelle dargestellt.

Bereich	Kriterium	Alternativen	Behandlung im D3TRAINER
Benutzerschnittstelle	Generizität der Falldarstellung	• Statisch • Generiert	Generierte Fallpräsentation aus dem D3 Expertensystem und Falldaten
	Interaktivität der Falldarstellung	• Automatische Fallbearbeitung • Teilweises Eingreifen des Studenten • Vollständige Steuerung der Fallbearbeitung durch den Studenten	Optionen: • Ganze Symptomatik sofort • Symptomatik in Aufwandsgruppen in fester Reihenfolge ohne Anforderung • Hauptsymptomatik sofort und weitere Untersuchungsergebnisse auf Anfrage • Hauptsymptomatik sofort und Anforderung von Einzelinformationen durch den Studenten

weiter auf nächster Seite

Bereich	Kriterium	Alternativen	Behandlung im D3TRAINER
	Detaillierungsgrad der Falldarstellung	• Grobe oder genaue Beschreibung der Symptome • Unterscheidung in abnorme und normale Symptome	• Detaillierte Beschreibung der Symptome • Hierarchische Ordnung der Symptome zu Oberbegriffen • Unterscheidung in abnorme und normale Daten optional
	Verwendete Medien	• Text • Bilder • Töne • Bewegtbilder	• Text • Bilder
	Darstellung von Texten	• Freie Formulierung von Texten • Schematisierte Darstellung von Texten	• Schematisierte Darstellung bei den Falldaten (generiert) • Freie Formulierung bei Erklärungs- und Hintergrundwissen (vorformuliert)
	Aktionsauswahl	• Offene Fragen • Semioffene Fragen • Geschlossene Fragen	• Semioffene Fragen über hierarchische Auswahl aus allen Daten • Optional MC-Auswahl aus Teilmenge
	Navigation	• Ereignisgesteuert (Menüzeile, Tastaturkürzel) • Geführt (Buttons)	• Hauptsächlich Navigation über Buttons • Selten benutzte Optionen in Menüs
	Orientierung	• Online-Manual • Kontextvisualisierung • Historie • Technische Bedienhilfen	• Kontextvisualisierung • Eingeschränkt: Historie
Studentenmodell	Modellierter Inhalt	• Präferenzen des Lerners • Wissen des Lerners	Präferenzen des Lerners
	Eingabe/ Generierung	• Explizit (Einzeleinstellungen, Stereotypenauswahl) • Implizit unterbrechend (Nachfrage, Fragebogen) • Implizit automatisch (Monitoring)	Explizit durch Einstellungen oder Stereotypenauswahl
	Repräsentation und Methodik	• Einzeleinstellungen • Stereotypen • Fehlerbibliotheken • Overlaymodelle	• Einzeleinstellungen • Stereotypen
	Nutzen	• Fallauswahl • Adaptive Oberfläche • Präsentation von informellen Wissen • Adaptive didaktische Strategie • Adaptive Kritik • Leistungskontrolle	• Festlegen der didaktischen Strategie, insbesondere des Feedbackverhaltens • Festlegen von Teilen der Oberflächengestaltung

weiter auf nächster Seite

Bereich	Kriterium	Alternativen	Behandlung im D3TRAINER
Didaktik	Steuerungsgrad der Lernaktivität	• Systemgesteuert • Lernergesteuert	Lernergesteuert
	Allgemeine Fähigkeiten	• Kognitive Fähigkeiten • Manuelle Fähigkeiten • Organisatorische Fähigkeiten • Soziale und kommunikative Fähigkeiten	Primär kognitive Fähigkeiten
	Kognitive Fähigkeiten	• Diagnoseauswahl • Therapieauswahl • Symptomerkennung • Untersuchungsauswahl • Kontinuierliches Monitoring	Alle kognitiven Aufgaben bis auf kontinuierliches Monitoring berücksichtigt, teilweise ausblendbar
	Feedback-algorithmus	• Generiert oder statisch • Einfacher Vergleich oder benutzerorientiert • Situations- oder ergebnisabhängig • Bewertungszeitpunkt	• Generiert • Benutzerorientiert, berücksichtigt: Ähnlichkeiten, Wichtigkeit, Sicherheitsfaktor • Situations- oder ergebnisbezogen • Bewertungszeitpunkt einstellbar: immer gleich, gebündelt oder erst am Fallende
	Erklärung	• Generiert oder vorformuliert • Allgemein oder situationsspezifisch • Automatisch oder auf Anfrage • Detaillierungsgrad	• Generierte situationsbezogene Erklärung • Vorformulierte allgemeine Erklärung • Auf Anfrage des Lerners • Hoher Detaillierungsgrad auf Basis des Wissensmodells
	Begründung des Lerners	• Eingabeform (offen, semioffen, geschlossen) • Begründung kritisierbar • Bezug auf Erklärung des Systems	• Student kann selbst aktiv Begründungen eingeben • Begründungen des Studenten werden vom System kritisiert
	Kontext-sensitive Entscheidungs-hilfen		• Bei jeder Studentenaktion ist eine kontextbezogene Entscheidungshilfe abrufbar • Spezieller Prüfungsmodus ohne Entscheidungshilfen wählbar
Wissens-modell	Form	• Seitenbasiert • Problemlösewissen • Simulationsmodell	Problemlösewissen
	Inhalt	• Fallwissen • Strukturelles Wissen • Strategisches Wissen • Begriffliches Wissen • Bildhaftes Wissen • Hintergrundwissen	Alle Wissensarten berücksichtigt
	Metawissen	• Metakognition (benutzerorientiert) • Problemlösungs-methode (systemorientiert)	Beide Arten von Metawissen abbildbar

weiter auf nächster Seite

Bereich	Kriterium	Alternativen	Behandlung im D3Trainer
	Wissens- akquisition	• Visuelle (durchgängig graphische Eingabe) oder textbasierte Programmierung • Autorensystem	• Spezielles Autorensystem • Graphische Wissenseingabe
	Mehrfach- verwendung	• Beratung • Kritik • Dokumentation • Suche nach ähnlichen Fällen • Statistik • Dokumenten-generierung	Alle Möglichkeiten der Mehrfachverwendung sind durch die Expertensystemshell D3 gegeben
	Evaluations- hilfen	• Linküberprüfung • Plausibilitätskontrolle • Falltesten	• Automatischer Wissensbasistest mit Fällen • Unterstützung von Plausibilitätskontrollen durch detaillierte Erklärung

Tabelle 2.1: Kriterien und Aspekte bei fallbasierten intelligenten Trainingssystemen, Alternativen und die Behandlung im D3Trainer.

Das erste Kriterium bei der **Benutzerschnittstelle** ist die Frage danach, wie die Falldarstellung erzeugt wird. Bei fallbasierten Trainingssystemen kann diese entweder statisch vorformuliert oder aus einem Wissensmodell generiert worden sein. Diese *Generizität der Falldarstellung* ist besonders wichtig bei Änderungen und Erweiterungen bei der Fallpräsentation. So muß bei ausformulierten Varianten jeder Fall einzeln geändert werden, während bei der generierten Darstellungen nur die Schablone geändert werden muß. Der D3Trainer basiert auf dem Expertensystemshellbaukasten D3 ([Puppe et al. 1996]) und generiert seine Falldarstellung aus dem Wissen von D3–Expertensystemen und zusätzlichen Falldaten.

Bei der *Interaktivität der Falldarstellung* handelt es sich um ein Kriterium, welches beschreibt, wie aktiv der Lerner in die Fallbearbeitung eingebunden ist. Bei einer automatischen Fallbearbeitung wird der Student vollständig durch den Fall geführt, ohne Eigeninitiative zeigen zu müssen. Man kann sich verschiedene Zwischenstufen vorstellen, in denen der Lerner teilweise in die Fallbearbeitung eingreifen kann bis hin zu einer alleinigen Steuerung des Ablaufes durch den Lerner. In dem D3Trainer sind verschiedene Optionen angeboten, die dem Lerner unterschiedlich viel Aktionsmöglichkeiten bereitstellen. Entweder wird ihm die gesamte Fallsymptomatik sofort präsentiert und er muß keine eigenen Tests (Untersuchungen) anfordern oder er kann den Fall in größeren Gruppen (Aufwandsgruppen) in einer bestimmten Reihenfolge abrufen. Eine weitere Option ist die Präsentation der Hauptsymptomatik und die Auswahl aller weiteren Untersuchungen durch den Lerner selbst. Die letzte Stufe bei der Interaktivität ist die Möglichkeit zur Auswahl einzelner Fragen nach anfänglicher

Präsentation der Hauptsymptomatik durch das System.

Die Darstellung der Fallsymptomatik in dem System kann in einem unterschiedlichem *Detaillierungsgrad* geschehen, so daß die Skala von allgemeinen Problembeschreibungen (z. B. Rückenschmerzen) bis hin zu detaillierten Aussagen (z. B. Druckschmerz im Lendenwirbelbereich) reicht. Auch kann zwischen abnormen und normalen Symptomen unterschieden werden, so daß der Lerner bereits in der Falldarstellung auf relevante Merkmale hingewiesen wird. Im D3TRAINER findet eine hierarchische Darstellung von elementaren Symptomen statt, mit denen auch das verwendete Expertensystem arbeitet. Die Symptome sind in einer auf– und zuklappbaren Hierarchie geordnet, so daß dem Lerner verschiedene Detaillierungsstufen angeboten werden. Die Merkmale im D3TRAINER sind in abnorme und normale Symptome eingeteilt, die optional farblich unterschieden werden können, wobei die normalen Symptome auch vollständig ausblendbar sind.

Bei der Falldarstellung kann auch nach den *verwendeten Medien* (Texte, Bilder, Töne, Bewegtbilder) unterschieden werden, wobei im D3TRAINER im Moment nur Texte und Bilder verwendet werden. Texte bilden den elementaren Bestandteil der Fallpräsentation und sind als zusätzliche informelle Erklärungseinheiten (z. B. Lehrbuchtexte) integriert. Die im D3TRAINER verwendeten Bilder haben entweder erläuternden Charakter (informelle Erklärung) oder werden für die Symptomerkennung benutzt, so daß der Lerner aus den Bildern Symptome erkennen und in die Fallpräsentation eingeben muß.

Die *Darstellung der Texte* kann von freien Formulierungen (Fließtext, zusammenhängende Sätze) bis hin zu schematischen Darstellungen (Schablonen) variieren. Die Falldaten im D3TRAINER werden aus dem Expertensystem generiert und in einer schematischen Darstellung (z. B. Fieber = 38 Grad) angezeigt, während bei der Darstellung von Erklärungs– und Hintergrundwissen eine freie Formulierung gewählt wird. Die Erklärungen werden auch aus dem Expertensystemwissen generiert und in ganze Sätze umgewandelt. Das Hintergrundwissen wurde durch die Fachexperten vollständig vorformuliert.

Nach der initialen Falldarstellung kann der Student verschiedene Aktionen auswählen, wobei diese *Aktionsauswahl* wiederum in unterschiedlicher Art und Weise geschehen kann. Neben den offenen (Freitext) und geschlossenen (z. B. MC–Auswahl, Zuordnungen) Fragen werden hier semioffene Fragen definiert. Bei diesem Fragetyp handelt es sich um eine geschlossene Frage, deren Antwortalternativen aber so umfangreich sind, daß sie einer offenen Frageform nahekommt. Der im D3TRAINER verwendete semioffene Fragetyp ist beispielsweise eine Auswahl aus einer hierarchischen Ordnung aller möglichen Antworten (Diagnosen oder Tests; im *RheumaTutor* durchschnittlich zwischen 100 und 200 Alternativen). Optional bietet der

D3Trainer aber auch eine MC–Auswahl aus ca. 15 Alternativen (Anzahl fest, aber beliebig) an.

Die *Navigation* in fallbasierten Trainingssystemen ist besonders bei benutzergesteuerten Varianten wichtig. Man unterscheidet hier zwischen ereignisgesteuerten und geführten Aktionen. Ereignisgesteuert bedeutet, daß der Lerner ein Ereignis (z. B. Diagnoseauswahl, Erklärung, neuen Fall wählen) herbeiführen möchte, wobei er meist ein Tastaturkürzel oder die Menüzeile benutzt. Bei geführten Aktionen werden dem Studenten situationsbezogen alle Möglichkeiten, beispielsweise mit Buttons, angeboten, aus denen er dann wählen kann. Die im D3Trainer vorherrschende Navigationsform ist die Führung des Lerners durch situationsbezogene Buttons. Selten benutzte Optionen (z. B. Neuer Fall) können durch den Benutzer ereignisgesteuert über die Menüzeile erreicht werden.

Als letzter Aspekt, der eng mit der Navigation zusammenhängt, sei hier die Unterstützung der *Orientierung* des Studenten in der Fallbearbeitung genannt. Gerade wenn der Lerner selbstgesteuert den Fall durchläuft, muß darauf geachtet werden, daß sich dieser immer bewußt ist, in welchem Kontext er sich gerade befindet. Grundsätzlich sind ein Online–Manual und sonstige technische Bedienhilfen (wie beispielsweise Tooltips) in diesem Zusammenhang wesentliche Hilfen, vor allem wenn sie eine situationsbezogene Unterstützung bieten können. Aber auch Kontextvisualisierungen, wie beispielsweise Flowcharts oder spezielle Hintergrundfarben der Fenster, helfen den Anwender, sich in der Fallumgebung zurecht zu finden. Auch eine Darstellung der bisherigen Historie, evtl. mit Rücksprungoption, kann von großem Vorteil sein. Der D3Trainer bietet in diesem Aspekt eine Kontextvisualisierung über Farben an, die aber noch durch aussagekräftige Hintergrundbilder bei den Fenstern der verschiedenen Bereiche verstärkt werden soll. Eine eingeschränkte Historie ist insoweit bereitgestellt, daß bei den möglichen Aktionen immer auf die bisher durchgeführten Aktionen dieser Art (z. B. Verdachtsdiagnose auswählen) verwiesen werden kann.

Der zweite Aspekt bei intelligenten Trainingssystemen ist das **Studentenmodell**, wie es seine Werte ermittelt und welchen Nutzen es für das Gesamtsystem bringt. Der *modellierte Inhalt* eines Studentenmodells ist die erste Unterscheidungsmöglichkeit verschiedener Ansätze. In einem Studentenmodell können sowohl Präferenzen des Benutzers wie auch ein Modell über sein vermutliches Wissen abgelegt werden. Der D3Trainer arbeitet ausschließlich mit Präferenzen und macht keine Aussagen über den Wissenstand. Die *Eingabe bzw. Generierung* des Studentenmodells kann entweder explizit oder implizit erfolgen. Bei einer expliziten Eingabe wird der Lerner aufgefordert, Einzeleinstellungen oder Stereotypen auszuwählen, die den späteren Ablauf beeinflussen. Eine implizite Eingabe bedeutet, daß der Lerner seine Wünsche und sein Wissen nicht direkt formuliert, sondern

die Entscheidung durch Fragebögen und Nachfragen durch das System (unterbrechend) oder durch Benutzermonitoring mit anschließender Analyse (automatisch) gefällt werden. Die Eingabe der Benutzerpräferenzen in dem D3TRAINER erfolgt explizit über die Wahl von Einzeleinstellungen oder durch die Entscheidung für einen Stereotypen, der eine Menge an Einzeleinstellungen repräsentiert. Abhängig von Inhalt und Eingabe des Studentenmodells ist die *Repräsentation und die Methodik*, die diesem zugrunde liegt. Alternativen dafür sind Einzeleinstellungen, Stereotypen, Fehlerbibliotheken und Overlaymodelle. Die beiden ersten Alternativen sind die im D3TRAINER verwendeten Methodiken.

Durch Inhalt, Eingabe und Repräsentation ist die Beschaffenheit des Studentenmodells festgelegt, der wichtigste Aspekt ist aber der *Nutzen*, den es für das Trainingssystem bringt. Mit einem Studentenmodell kann beispielsweise die Fallauswahl automatisch unterstützt werden, so daß der Student immer einen Fall präsentiert bekommt, der für seinen momentanen Anforderungen und Wissensstand geeignet ist. Auch die Oberfläche kann entsprechend des Studentenmodells adaptiert werden, so daß beispielsweise die Navigationsunterstützung auf den Benutzer abgestimmt wird. Auch bei der Präsentation von informellem Wissen kann das Studentenmodell genutzt werden, um den Studenten immer genau das Wissen anzubieten, welches seinen Problemen und seinem Wissenstand entspricht. Neben der adaptiven Oberfläche kann auch die Lernstrategie an die Bedürfnisse des Lerners adaptiert werden, wenn z. B. ein Lerntyp durch das Studentenmodell bestimmt wird. Auch die Leistungskontrolle über mehrere Fälle hinweg, kann mittels eines Studentenmodells besser generiert und genutzt werden. Durch die Einzeleinstellungen des D3TRAINERS wird die didaktische Strategie festgelegt, besonders aber das Feedbackverhalten des Systems. Aber auch die Oberflächengestaltung läßt sich durch diese Einstellungen parametrisieren.

Die **Didaktik**, die in einem intelligenten fallbasierten Trainingssystem benutzt wird, wird durch verschiedene Entscheidungen festgelegt. Die elementare Entscheidung besteht in der Festlegung des *Steuerungsgrades der Lernaktivität*, d. h. wer ist der aktive Teil des Lerndialoges. Die meisten Systeme gehören zu einem der Extreme ‚lernergesteuert' oder ‚systemgesteuert', wobei es auch viele Stufen dazwischen gibt. Der D3TRAINER ist weitgehend lernergesteuert, bietet aber einige didaktische Varianten an, bei denen das System den Benutzer stärker führt. Zu dem Punkt Didaktik gehört auch die Aussage, welche *allgemeinen Fähigkeiten* das System dem Lerner vermitteln will. Neben den kognitiven Fähigkeiten (z. B. Schlußfolgern) gehören manuelle und organisatorische, sowie soziale und kommunikative Fähigkeiten zu einer vollständigen Abdeckung des Problembereichs. In dem D3TRAINER werden primär die kognitiven Aufgaben eingeübt, während andere Fähigkeiten aus den Bereichen Organisation oder Kom-

munikation eher sekundär über Hintergrundwissen und Nebenaufgaben berücksichtigt werden.

Kognitive Aufgaben bei der Diagnostik betreffen Auswahl von Diagnosen und Therapien, Symptomerkennung, sowie die Auswahl weiterer Untersuchungen (Tests). Bei allen Aufgaben kann unterschieden werden, ob es eine einmalige Aktion ist oder ob ein kontinuierliches Monitoring, beispielsweise Therapieüberwachung, vorliegt. In dem D3TRAINER sind alle kognitiven Aufgaben ohne kontinuierliches Monitoring abgebildet, wobei die Aufgaben auch teilweise ausblendbar sind. Soll der Student bei der Durchführung dieser Aufgaben durch das System kritisiert werden, so muß der *Feedbackalgorithmus* näher betrachtet werden. Dieser kann wieder entweder generiert oder statisch orientiert sein, wobei die Benutzerlösung entweder mit der Systemlösung durch einen einfachen Vergleich bewertet oder eine benutzerorientierte Bewertung genutzt wird. Bei Aktionen, die auf eine momentane Fallsituation bezogen sind, wie beispielsweise die frühe Angabe einer Verdachtsdiagnose, kann der Algorithmus entweder mit der letztlich korrekten Diagnose (ergebnisorientiert) oder mit der aktuellen Diagnose vergleichen, die in dieser Fallsituation (situationsorientiert) angebracht ist. Auch der Bewertungszeitpunkt muß bei der Beschreibung des Feedbackalgorithmus betrachtet werden. Die Bewertung kann entweder immer gleich, erst auf Anforderung oder erst nach der Fallbearbeitung gegeben werden. Der D3TRAINER verwendet eine generierte, stark benutzerorientierte Bewertung, die Ähnlichkeiten, Wichtigkeit und Sicherheitsfaktoren berücksichtigt. Sie ist durch Einstellungen entweder situations– oder ergebnisorientiert nutzbar und wird normalerweise nur auf Anfrage des Lerners präsentiert. In manchen Szenarien kann sie aber auch ganz ausgeschaltet werden und erscheint erst zu Ende der Fallbearbeitung, bei einem stark geführten Dialog werden einige Bewertungen automatisch präsentiert.

Eng mit dem Feedbackalgorithmus verbunden ist die *Erklärung* des Systems, die ebenfalls generiert oder vorformuliert sein kann. Auch sie kann entweder allgemeiner Natur oder situationsspezifisch sein und automatisch oder auf Anfrage des Benutzers erscheinen. Ein wichtiger Punkt ist hierbei auch der Detaillierungsgrad der Erklärung und die Tatsache, ob die Erklärung eventuell in verschiedenen Stufen (z. B. Hinweis, deutlicher Hinweis, Lösung) abrufbar ist. Im D3TRAINER liegt eine generierte situationsbezogene Erklärung sowie eine frei formulierbare allgemeine Erklärung vor, die nur auf Anfrage des Studenten erscheint. Durch die Nutzung des Wissensmodells kann ein hoher Detaillierungsgrad erreicht werden, der allerdings nur selten mehrstufig angeboten wird. Ein Beispiel dafür ist die Erklärung zur Bewertung der Diagnosenauswahl, bei der in der ersten Stufe nur die Einordnung der gewählten Diagnosen in falsch oder richtig erkannt angeboten wird, während bei der zweiten Stufe die Hinweise aufgerufen werden können, die für bzw. gegen diese Diagnosen sprechen.

Wenn der Lerner Aktionen in dem Trainingssystem durchführen kann, ist es auch interessant, ob er für diese Aktionen *Begründungen* angeben kann. Auch hier kann bei der Eingabeform zwischen offenen, semioffenen und geschlossenen Frageformen unterschieden werden. Wird die angegebene Begründung durch das System kritisiert und bewertet, sollte dies in Bezug auf die Erklärung des Systems stattfinden. Grundsätzlich sind hier wieder alle Anmerkungen zu dem Feedbackalgorithmus (siehe oben) anwendbar. Der D3TRAINER ermöglicht dem Studenten, für alle kognitiven Aufgaben außer der Symptomerkennung, aktiv Begründungen einzugeben. Diese Begründungen werden durch einen benutzerorientierten Vergleich mit den Erklärungen des Systems bewertet und diese Bewertung wird dem Studenten wiederum erklärt.

Der letzte Punkt bei der Beschreibung der Didaktik eines Systems ist die Bereitstellung *kontextsensitiver Entscheidungshilfen*, also Hilfen, die dem Studenten in einer bestimmten Situation bei der Auswahl der richtigen Aktion unterstützen. In dem D3TRAINER ist zu jeder Aktion eine kontextbezogene Entscheidungshilfe abrufbar, die die korrekte Lösung für diese Situation anzeigt. Allerdings wird auch ein Modus angeboten, der diese Entscheidungshilfen vollständig ausblendet, so daß eine Prüfungssituation simuliert werden kann.

In intelligenten Trainingssystemen wird ein **Wissensmodell** benutzt, in dem das zu vermittelnde Wissen in verschiedenen *Formen* repräsentiert wird. Dabei gibt es seitenbasierte Systeme, die Dokumentseiten als elementare Wissenselemente benutzen, Systeme mit eigenem Problemlösewissen und Simulationssysteme, die ein ausführbares Modell der Lernwelt bereitstellen. Bei dem D3TRAINER handelt es sich um ein System mit Problemlösewissen, welches auf einem Expertensystem aufbaut. Der *Inhalt* des Wissens eines fallbasierten Trainingssystems unterteilt sich in Fallwissen, strukturelles Wissen (z. B. Attribute einer Lösung), strategisches Wissen (z. B. *Welche Untersuchung ist die nächste?*), begriffliches Wissen (z. B. *Wie errechnet man den ‚body mass index'?*), bildhaftes Wissen (z. B. *Wie sieht ein Ekzem aus?*) und Hintergrundwissen (z. B. allgemeine Hinweise bei Arthritis). Der D3TRAINER erlaubt den Autoren die Eingabe aller Wissensarten, so daß diese dem Studenten auch vermittelt werden können.

Neben dem eben genannten inhaltlichem Wissen gibt es in Trainingssystemen auch *Metawissen*, also Wissen wie man mit Wissen umgehen muß. Dabei kann primär in Metakognition und Problemlösungsmethode unterschieden werden. Metakognition ist das Wissen, welches beim Lerner für die grundsätzliche Herangehensweise an ein Problem zuständig ist, während die Problemlösungsmethode eine rein technisch orientierte Strategie des Systems ist. Beide Arten von Metawissen sind in dem D3TRAINER abbildbar.

Bei der Eingabe des Wissens in das System, der sogenannten *Wissensak-*

quisition, kann unterschieden werden, ob das Wissen in einer textbasierten Programmiersprache formuliert werden muß oder über eine durchgängig graphische Eingabe (visuelle Programmierung) modelliert wird. In vielen Bereichen kann ein *Autorensystem* zur Eingabe des Wissens benutzt werden, wobei auch diese unterschiedlich komfortabel und passend für den Fachexperten sind. Der D3TRAINER stellt ein spezielles und damit sehr komfortables Autorenwerkzeug dar, in dem die graphische Wisseneingabe des Expertensystems D3 vollständig genutzt wird. Oft ist der Aufbau von Wissensmodellen mit eigener Problemlösefähigkeit mit einem höheren Basisaufwand verbunden, der sich aber später bei Wartung und Erweiterungen auszahlt. Ein weiterer Faktor, der für ein solches ausführbares Wissensmodell spricht, ist die Möglichkeit der *Mehrfachnutzung*. Gerade im Bereich der Diagnostik, insbesondere in einer medizinischen Anwendung, kann das Wissen auch für Beratung (Konsultation), Kritik, Falldokumentation, Suche nach ähnlichen Fällen, Statistik, sowie der Dokumentengenerierung (z. B. Arztbrief) genutzt werden. All diese Optionen sind bereits in dem Expertensystem D3 bereitgestellt oder werden geplant, weshalb ein Wissensmodell des D3TRAINERS in vielfacher Weise in anderen Szenarien angewendet werden kann.

Die bereits angesprochenen Wartung und Erweiterung des Trainingssystems kann zu Problemen bei der Konsistenzhaltung des Wissensmodells führen. Eine notwendige Erweiterung bei diesen Prozessen sind *Evaluationshilfen*, die bei einer Änderung des Wissensmodells aufgerufen werden können. Beispiele für solche Hilfen sind Linküberprüfung, Plausibilitätskontrollen oder das Testen mit Fällen. In dem D3TRAINER wird ein automatisches Wissensbasistesten durch Fälle angeboten. Die Fälle werden in dem Expertensystem mit der aktuellen Wissensbasis durchgespielt und die tatsächlich hergeleitet Lösung mit der abgespeicherten korrekten Lösung verglichen, so daß der Fachexperten sofort einsehen kann, ob eine Änderung in der Wissensbasis negative Auswirkungen auf die Lösungen der Fälle hat. Plausibilitätskontrollen werden zwar nicht automatisch durchgeführt, aber durch eine detaillierte Erklärung stark unterstützt.

Der in den Expertensystembaukasten D3 integrierte D3TRAINER bietet viele Ansatzpunkte für die Generierung didaktisch variabler Trainingssysteme. Die Anwendbarkeit in medizinischen Bereichen wurde in mehreren Applikationen erwiesen, wobei sich das System auch für andere Bereiche der diagnostischen Problemlösung eignet.

3

Konzept zur Generierung fallbasierter Trainingssysteme

Die Erstellung von fallbasierten Trainingssystemen, die in der problemorientierten Lehre von großen Nutzen sein können, ist für den Fachexperten ohne Programmierkenntnisse keine leichte Aufgabe [PUPPE and REINHARDT 1995]. Intelligente Tutorsysteme, die auf einem Expertensystem basieren, bieten die besten Möglichkeiten für die Umsetzung solcher komplexen, interaktiven Lernumgebungen. Folglich sollte es den Fachexperten ermöglicht werden, selbständig ihr Domänenwissen zu formulieren und daraus ein Lernsystem zu generieren. In diesem Kapitel wird der Expertensystembaukasten D3 vorgestellt, dessen tutorielle Komponente D3TRAINER aus der vorhandenen Wissensbasis und einer Fallsammlung ein fallbasiertes Trainingssystem zum diagnostischen Problemlösen generiert (siehe auch [REINHARDT and SCHEWE 1995] und [REINHARDT 1997]). In Bezug auf die verschiedenen Möglichkeiten zur Kritik und die didaktischen Variationen, die in dem letzten Kapitel präsentiert wurden, wird anschließend deren Umsetzung in dem D3TRAINER vorgestellt.

3.1 Expertensystemshellbaukasten D3

Der Expertensystemshellbaukasten D3 [PUPPE et al. 1996] ist ein Werkzeug, mit dem eine Problemlöseumgebung für diagnostische Probleme konfiguriert werden kann. Man unterteilt hier grob in Wissenserwerb, also der Umgebung, in der ein Fachexperte sein Bereichswissen formuliert, und der Wissensnutzung, in der das Wissen auf verschiedene Weise genutzt werden kann (siehe Abb. 3.1).

Die Wisseneingabe durch den Bereichsspezialisten erfolgt rein graphisch und beruht auf den Basiseditoren Formular, Hierarchie und Tabelle (siehe

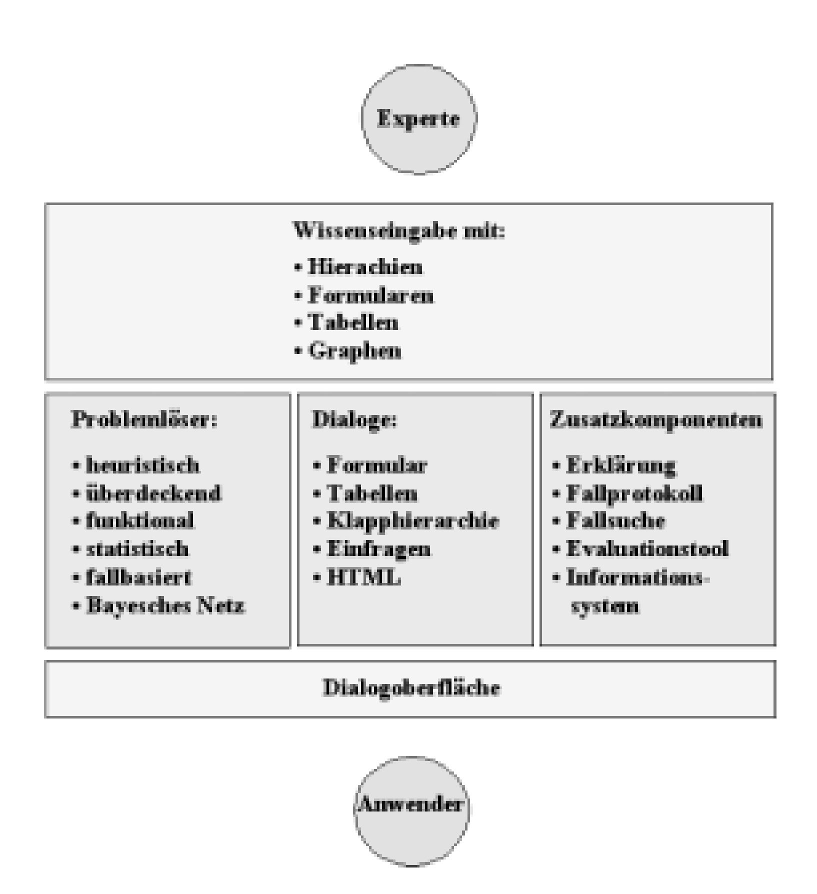


Abbildung 3.1: Grundlegende Komponenten in dem Expertensystembaukasten D3: Die Wissenseingabe wird durch den Fachbereichsexperten durchgeführt. Der Anwender kann über verschiedene Dialog Falldaten eingeben, die dann in den vorhandenen Problemlösern verarbeitet werden. Zusätzlich sind Komponenten für mehrere Zusatzaufgaben integriert, die für den komfortablen Einsatz in der Praxis benötigt werden.

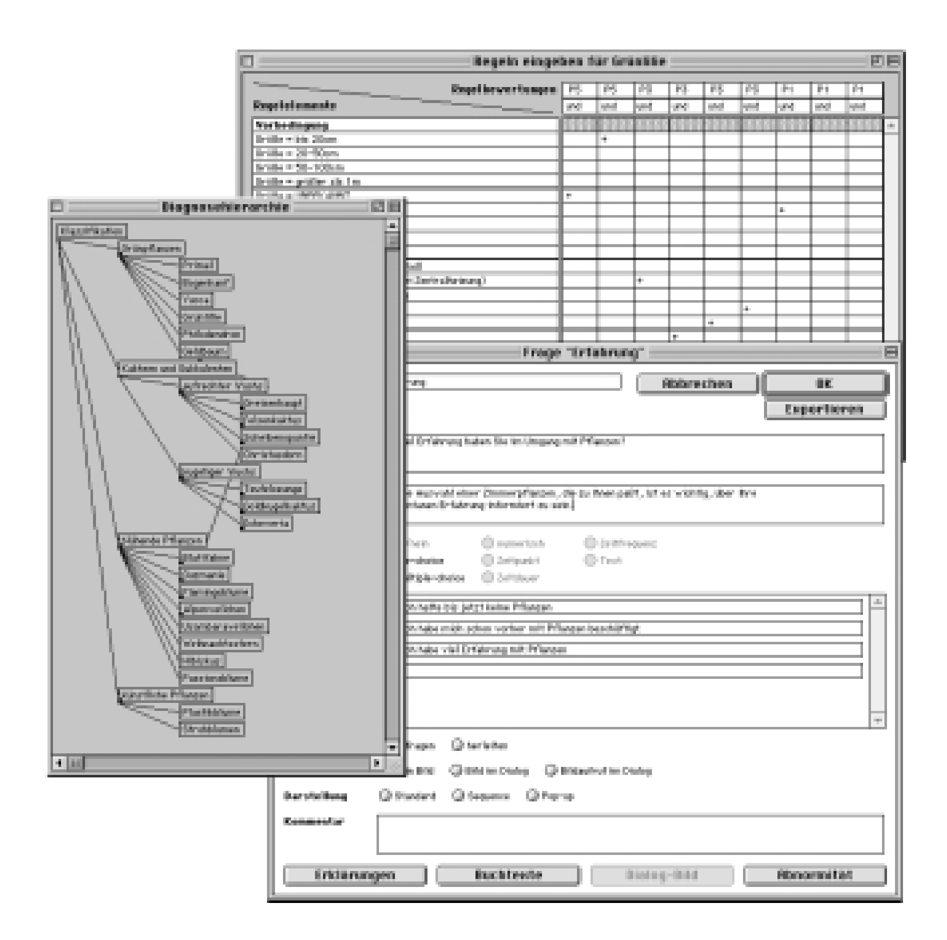


Abbildung 3.2: Beispiele für die Basiseditoren in der Expertensystemshell D3: Hierarchie zur Eingabe von Diagnosen, Formular zur Spezifikation eines Symptoms und Tabelle zur Eingabe von einfachen Regeln am Beispiel einer Wissensbasis zur Beratung beim Zimmerpflanzenkauf.

Abb. 3.2). Sie wird vollständig über Menüs gesteuert, so daß der Experte keine Programmierkenntnisse besitzen muß und die Möglichkeit einer syntaktisch fehlerhaften Eingabe gering gehalten wird.

Um die Wisseneingabe von D3 zu erklären, muß die Wissensrepräsentation in groben Zügen eingeführt werden. Die beiden Basisobjekte sind Symptome und Diagnosen, die über Regeln miteinander verknüpft werden. Die Regeln können in Regeln zur Diagnostik, also zur Herleitung eines Objektes (z. B. Diagnose), und zur Dialogführung (z. B. Stellen einer Folgefrage) eingeteilt werden.

Bei den Diagnosen unterscheidet man in Grobdiagnosen, Feindiagnosen und Kontextdiagnosen, die nur in dem diagnostischen Mittelbau ihre Funktion haben und bei der Ergebnisausgabe nicht erscheinen. Die Symptome sind in Frageklassenoberbegriffe, Frageklassen, Fragen und Sym-

ptomabstraktionen eingeteilt. Frageklassenoberbegriffe und Frageklassen dienen nur der Ordnung von Fragen. Die Fragen des Dialoges müssen durch den späteren Benutzer eingegeben werden und Symptomabstraktionen präsentieren wiederum den diagnostischen Mittelbau (Abb. 3.3 verdeutlicht diese Aufteilung). Eine Frageklasse besitzt mehrere Standardfragen, die grundsätzlich immer gestellt werden. Die Beantwortung dieser Fragen beeinflußt dann das Stellen weiterer, sogenannter Folgefragen. Die nächste Frageklasse wird entweder direkt durch eine Regel ausgesucht oder sie dient der Überprüfung einer Verdachtsdiagnose, die das System gestellt hat. Alle in Frage kommenden Frageklassen werden durch eine Kosten–Nutzen–Abschätzung geordnet und die beste davon als nächstes gestellt. Auf weitere Details der Dialogsteuerung wird hier nicht eingegangen, sondern auf [PUPPE et al. 1996] verwiesen.

In D3 können folgende Problemlösungsmethoden verwendet werden:

- Entscheidungsbaum:

  Der Entscheidungsbaum wird graphisch in einer Hierarchie angegeben, in der die entsprechenden Abfragen und Folgefragen bis hin zu den Diagnosen über Pop–Ups angehängt werden.

- Heuristische Regeln:

  Die heuristischen Regeln können über verschiedene Tabellen und Formulare eingegeben werden, wobei sie neben einer mehrstufigen Vorbedingung mit den Operatoren *und, oder, n–aus–m*[1] und *nicht*, komplexe Ausnahmen und Kontextdiagnosen sowie eine Einstufung haben, wie sehr sie für oder gegen das entsprechende Diagnostikobjekt (z. B. Diagnose) sprechen. Abb. 3.4 stellt eine heuristische Regel in einem Formular dar.

- Überdeckende Regeln: Die Eingabe von überdeckenden Regeln entspricht ungefähr der von heuristischen Regeln, nur zeigen sie von den Diagnostikobjekten auf die Symptome, die sie erklären können. Ihre Erklärungsstärke wird ebenfalls gewichtet und mit dem grundsätzlichen Gewicht des Symptoms verrechnet.

[1] *n–aus–m* bedeutet, daß mindestens *n* und höchstens *m* der angegebenen Bedingen erfüllt sein müssen.

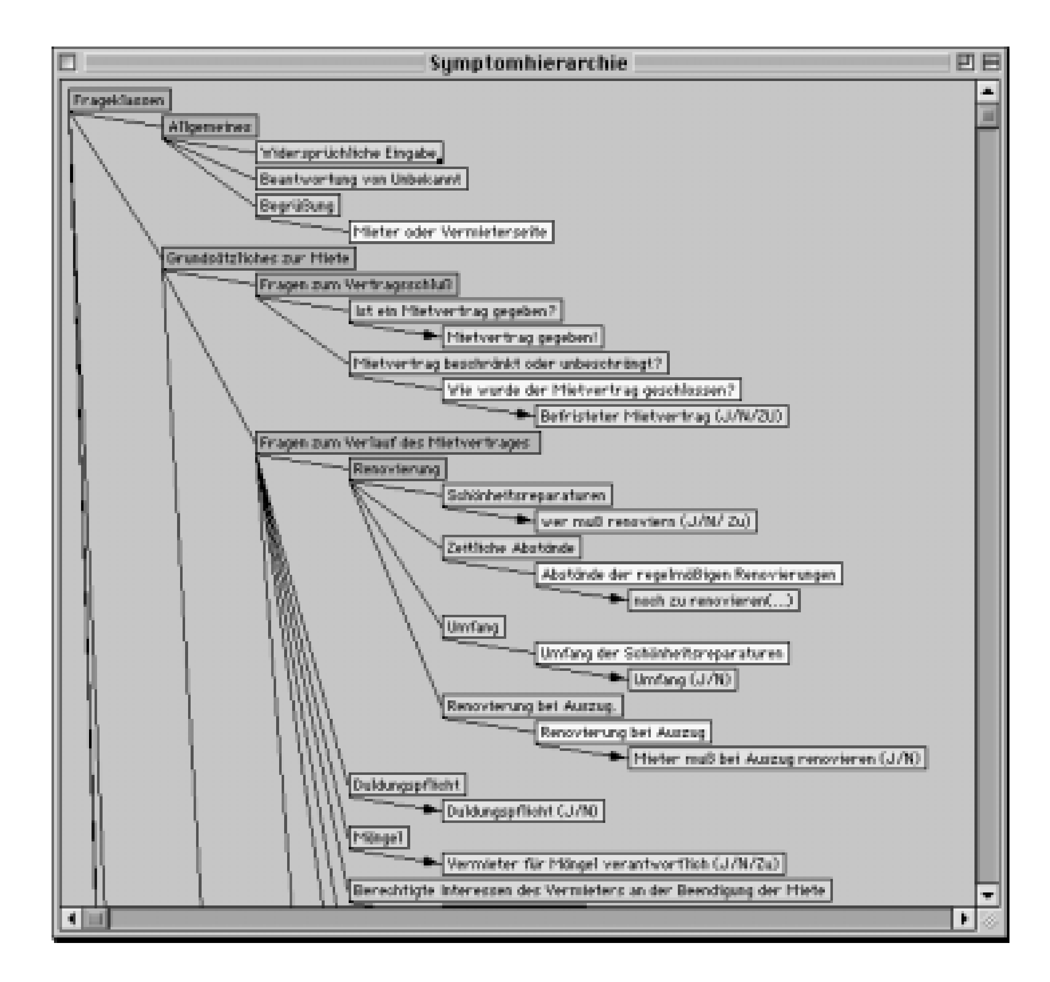


Abbildung 3.3: Beispiel für die Symptomhierarchie einer Wissensbasis. In dem hier dargestellten Beispiel zur juristischen Unterstützung bei einer Mietkündigung dienen die Frageklassen (z. B. *Fragen zum Vertragsschluß*, *Fragen zum Verlauf des Mietvertrages*) zur Ordnung der einzelnen Fragen (z. B. *Ist ein Mietvertrag gegeben?*). Es wird zwischen Fragen (z. B. *Gibt es Mängel?*) und deren Folgefragen (z. B. *Wer ist für diese Mängel verantwortlich?*) unterschieden, ebenso wie es Symptomabstraktionen (z. B. *Renovierung bei Auszug*) gibt, die aus den Fragen hergeleitet werden und nicht von dem Benutzer erfragt werden müssen.

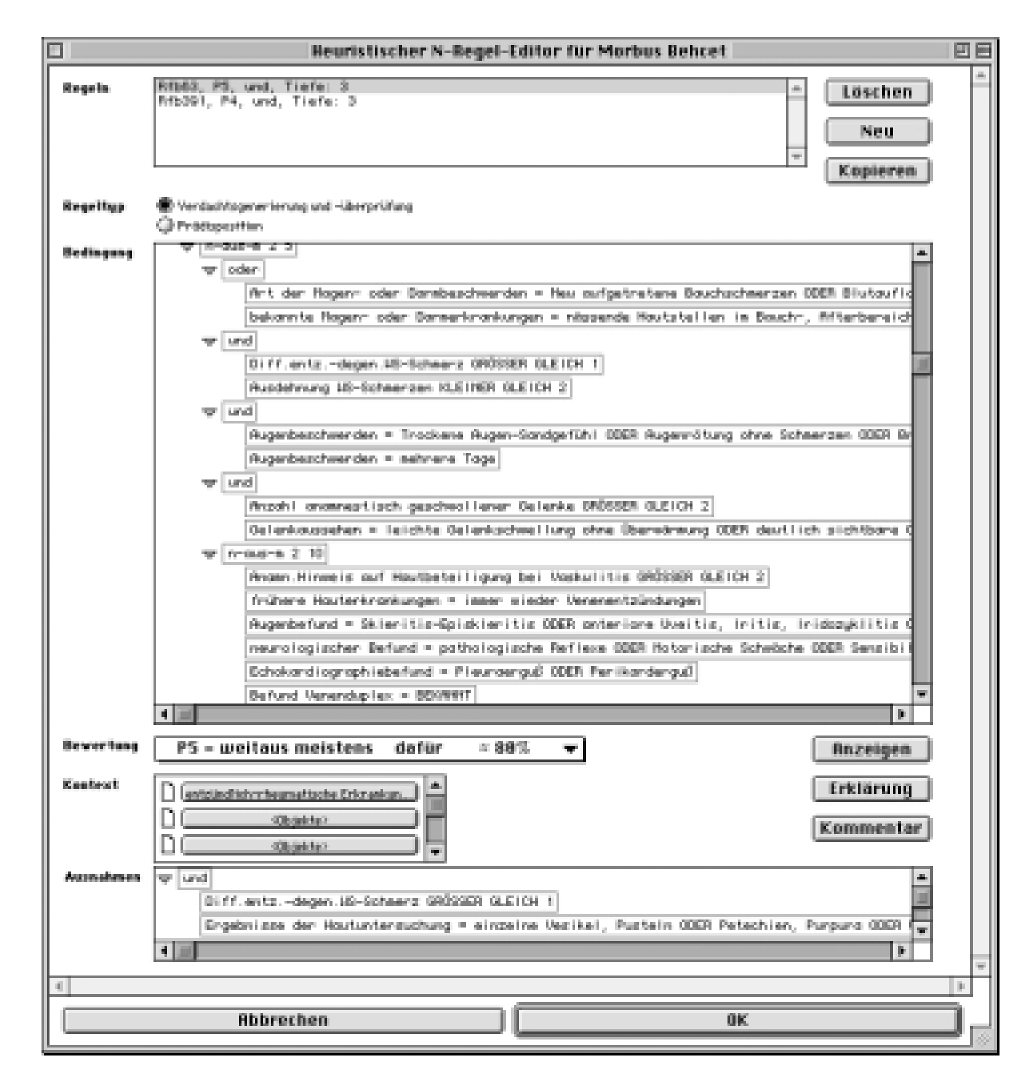


Abbildung 3.4: Formular zur Eingabe einer komplexen Regel zu der rheumatischen Erkrankung *Morbus Behcet*: Die mehrstufige Bedingung mit den Operatoren *und*, *oder* und *n–aus–m* entspricht einer internationalen Richtlinie.

- Funktionales Modell:

  Ein funktionales Modell wird in einem Graphen eingegeben, in dem Komponenten und Materialien festgelegt werden, deren Verhalten dann wiederum über Regeln beschrieben wird. Näheres zu dem funktionalen Problemlöser wird in [KOHLERT 1995] beschrieben.

- Statistischer Problemlöser und Bayessche Netze:

  Der statistische Problemlöser benötigt zum Ablauf eine Fallsammlung, aus der er die entsprechenden Wahrscheinlichkeiten ziehen kann. Der statistische Problemlöser und der Problemlöser mit Bayessche Netzen leiten ihre Lösungen durch die Berechnung von Wahrscheinlichkeiten her (mehr in [TOLKSDORF 1998]).

- Fallbasierter Problemlöser:

  Neben einer großen Fallsammlung, die für den Vergleich eines Falles mit den vorhanden Fällen benutzt wird, benötigt man beim fallbasierten Problemlöser (siehe [GOOS 1995]) Ähnlichkeitswissen, welches aussagt, wie ähnlich zwei Symptomausprägungen zueinander sind. Als Lösung des aktuellen Falles wird dann die Lösung des ähnlichsten Falles vorgeschlagen.

Die verschiedenen Problemlöser in D3 können ein– und ausgeschaltet werden, wobei die einzelnen Lösungen nach Ende der Falleingabe getrennt voneinander dargestellt werden (näheres dazu in [POECK 1995]). In neueren Ansätzen wird nun versucht, die Problemlöser verknüpft zu nutzen, um zu einer besseren Lösung zu gelangen [IGLEZAKIS 1999].

Für die Eingabe des Falles durch den Benutzer stehen in D3 mehrere Dialoge bereit (siehe Abb. 3.5), auf die nicht näher eingegangen wird. Sie sind auf die verschiedenen Wünsche in unterschiedlichen Bereichen anpaßbar, wobei der WWW–Dialog die Schnittstelle für eine Nutzung des Expertensystems als Server im World Wide Web darstellt [BERBERICH and BAMBERGER 1998].

Diagnose– bzw. Informationssysteme können nur dann erfolgreich in einer Problemlöseumgebung eingesetzt werden, wenn sie eine Vielzahl an zusätzlichen Aufgaben erfüllen oder sich in eine gegebene Struktur (z. B. elektronische Patientenakte mit Laborerfassungssystem) einbetten lassen. Der Expertensystembaukasten D3 stellt mehrere Optionen zur Verfügung, die nicht direkt zu dem Problemlöseprozeß gehören, die den Einsatz des Systems für den Anwender aber sinnvoller machen.

Die erste Zusatzkomponente, auf die hier eingegangen wird, ist allerdings noch sehr stark in den Problemlöseprozeß integriert. Die Erklärkomponente macht die in D3 ablaufenden Entscheidungsprozeße für den Benutzer durchsichtig. Sie baut auf einer hypertextartigen Verknüpfungsstruktur

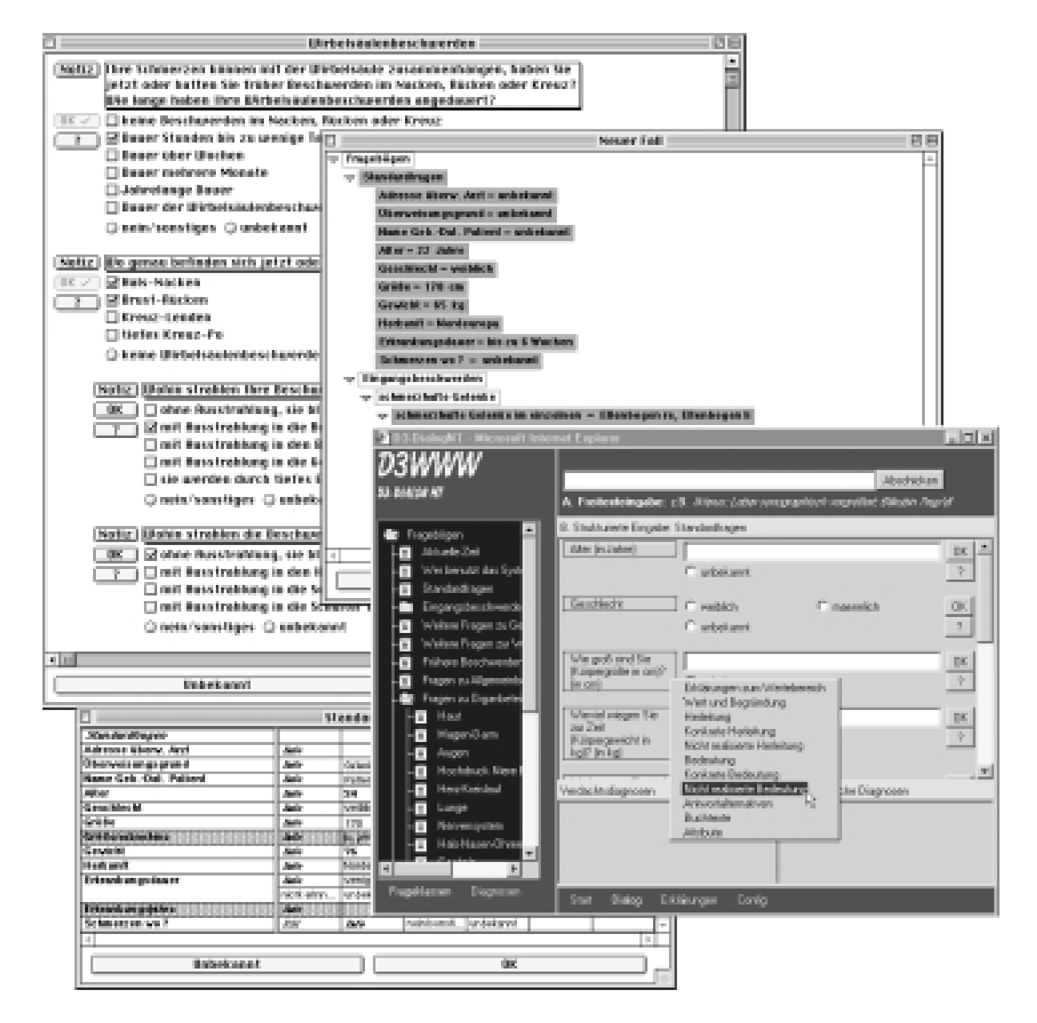


Abbildung 3.5: Unterschiedliche Dialoge in D3: Im Hintergrund links ist der Standarddialog, der auf Formularen beruht. Rechts oben ist der Klappdialog, der dem Benutzer ermöglicht, seine Eingaben in einer aufklappbaren Hierarchie zu machen und somit immer einen Überblick über den gesamten Fall zu haben. Unten links ist ein Fenster des Tabellendialogs gezeigt, der eine schnelle Eingabe für den geübten Benutzer der Wissensbasis darstellt, da die meisten Sonderinformationen nicht angezeigt werden. Im Vordergrund rechts ist eine Seite des webbasierten Dialoges, der in einem Standardbrowser abgerufen werden kann.

auf, bei der man sich zu jedem Begriff zusätzliche Informationen anzeigen lassen kann. In Abb. 3.6 wird ein Beispiel für die Erklärung einer heuristischen Diagnose gegeben. Die Erklärung erfüllt aber auch andere Aufgaben:

- Fallbezogene, dynamische Erklärungen zu allen Wissensarten: Dazu gehören beispielsweise Erklärungen zum Stand des Falldialoges (z. B. Warum wird welche Frageklasse gestellt?) und Erklärungen zu allen Problemlösern bezüglich aller Diagnostikobjekte, also zum Beispiel:
  - Welche Diagnosen wurden gefunden?
  - Welche Symptome sprechen für die Diagnose?
  - Welche Regeln haben gefeuert?
  - Warum haben andere Regeln nicht gefeuert?
- Fallunabhängiges, formales Wissen zu allen Wissensarten: Zu diesen Erklärungen zählen die Darstellung von Eigenschaften der Basisobjekte Symptome und Diagnosen, sowie Erklärungen bei der Falleingabe, um den Benutzer bei der richtigen Symptomeingabe zu unterstützen.

Die Erklärung ist ein übergreifendes Konzept, welches sich durch alle Komponenten zieht, bei denen Erklärungsbedarf besteht.

Das Erstellen eines Fallprotokolls aus den eingegebenen Falldaten ist eine wichtige Funktionalität, wenn das ausgedruckte Protokoll in Inhalt und äußerer Form genau den Wünschen des Anwenders entspricht. In dem Expertensystembaukasten D3 existiert ein hochgradig parametrisierbares Textgenerierungstool, welches dem Anwender hilft, Fallprotokolle in verschiedener Form zu generieren. Die einfachste Form ist ein Dokument, in dem eingegebene Daten und die Systemdiagnosen vermerkt sind. Die entsprechend aufwendigere Möglichkeit besteht darin, aus den Falldaten und einer vorgegebenen Schablone komplexe Textdokumente zu gewinnen. Ein Beispiel für eine solche Anwendung wäre das Generieren eines Arztbriefes aus einem Fall und einer Schablone, so daß der Arzt einfach einen Bericht an den Hausarzt schicken kann, in dem die richtigen Anreden, die wichtigen Symptome und die gefundenen Diagnosen vermerkt sind.

Die Verwaltung der Fälle ist ebenfalls eine Aufgabe, die durch D3 gut unterstützt wird. So kann in einer Sammlung von Fällen nach verschiedenen Merkmalen gesucht werden (siehe Abb. 3.7). Durch diese Funktionalität können sehr schnell spezifische Fälle gefunden werden. Besonders bei der Verwendung in der Lehre können so reale Fälle mit einer speziellen Charakteristik komfortabel gesucht werden.

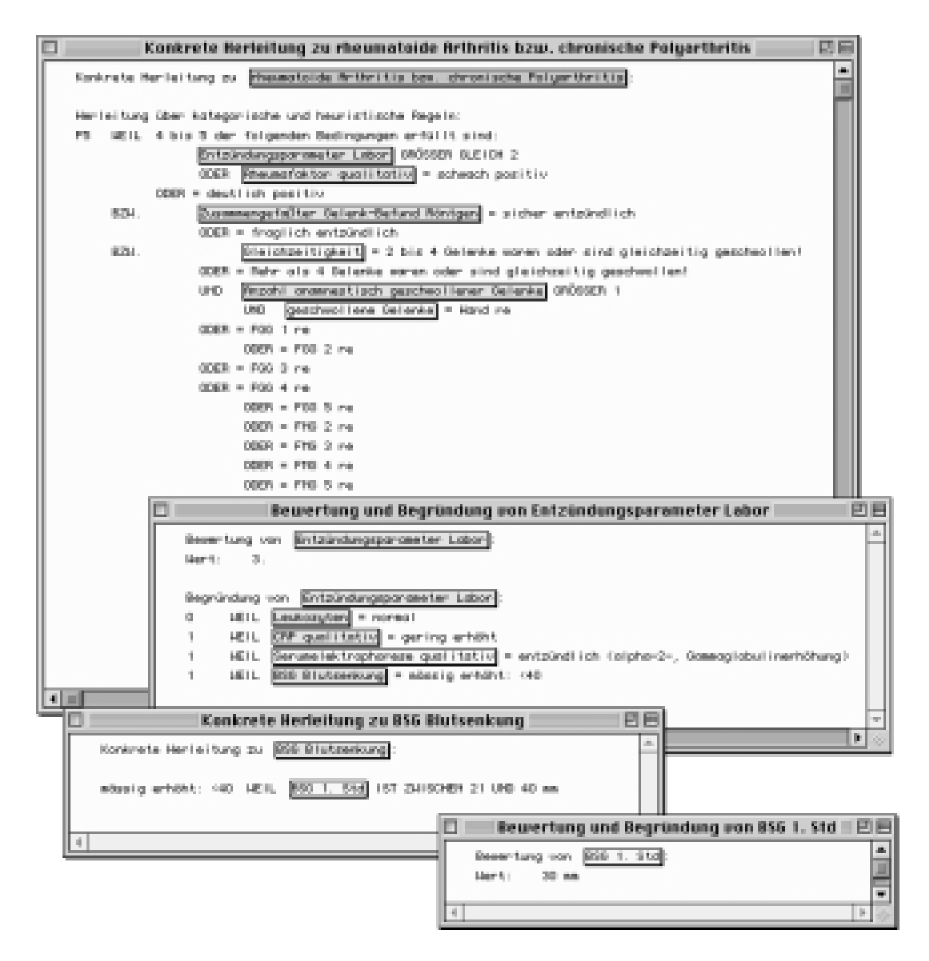


Abbildung 3.6: Gezeigt wird die aktuelle Herleitung zur Diagnose *rheumatoiden Arthritis*, über die Symptominterpretation *Entzündungsparameter Labor* und deren Herleitung über die Symptominterpretation *BSG Blutsenkung*. Ganz unten die Herleitung dieser Symptomabstraktion aus dem Rohdatum *BSG 1. Std.* Alle Herleitungen wurden über Menüs aus der anderen Herleitung aufgerufen.

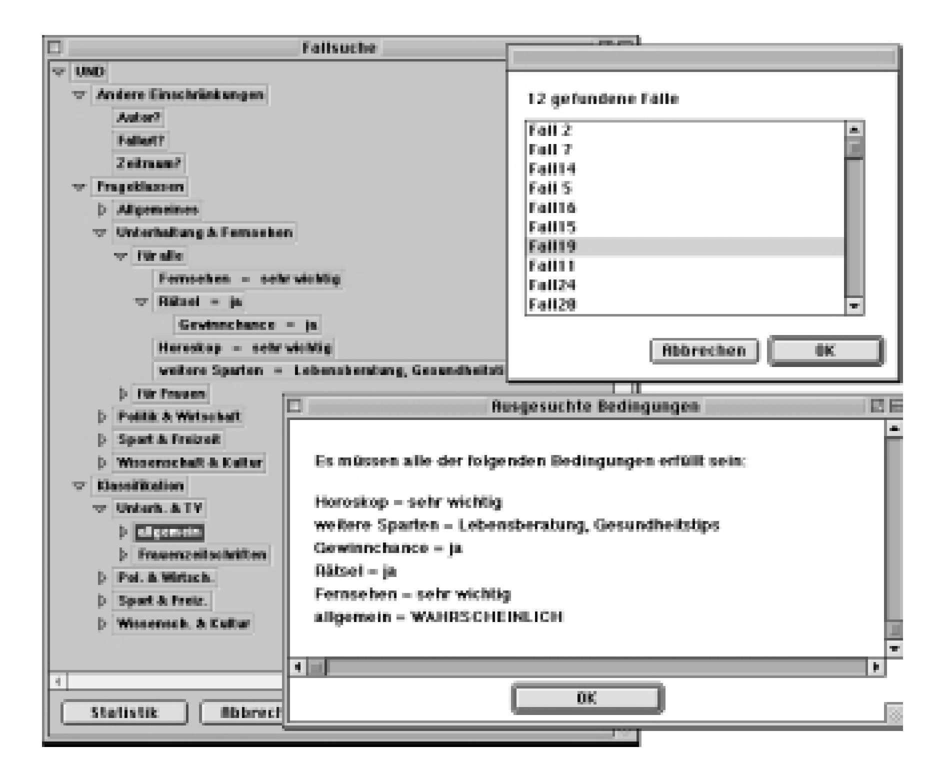


Abbildung 3.7: In einer Sammlung von Fällen kann gezielt nach Fällen mit einer bestimmen Symptomatik und Grobdiagnosen gesucht werden. Weitere Einschränkungen sind auch Autor des Falles, Zeitraum und Fallart (echter Fall, Musterfall, Phantasiefall). Das Beispiel aus einer Wissensbasis zur Auswahl einer passenden Zeitschrift zeigt die Sucheingabe in einer dynamischen Hierarchie, die zusammenfassende Anzeige der Suchkriterien und die Anzeige der gefunden Fälle, deren Inhalte dann dargestellt und durchgespielt werden können.

Ebenso kann eine Fallsammlung auch benutzt werden, um eine Wissensbasis zu evaluieren, also zu überprüfen, ob die korrekten Diagnosen durch das System gefunden werden. In einem Fall werden neben den Symptomen und Diagnosen des Systems ebenfalls die korrekten Diagnosen gespeichert, die ein menschlicher Experte eingegeben hat. Wurde die Wissensbasis verändert oder möchte man eine Leistungsabschätzung vornehmen, so kann mittels der Fallsammlung ein Test der Wissensbasis erfolgen. Dabei werden alle Fälle in den beliebigen (vorhandenen) Problemlösern durchgespielt und mit den realen Lösungen verglichen. Für jeden Fall wird nun eine Bewertung pro Problemlöser erhoben und diese am Abschluß über alle Fälle gemittelt. Auch diese Funktionalität des Expertensystembaukastens D3 ist für die Verwendung in der problemorientierten Lehre von großem Nutzen, da so alle Fälle nochmals auf ihre Korrektheit in D3 überprüft werden können, nachdem Änderungen in der Wissensbasis vorgenommen wurden.

Als letzte der Komponenten sei hier das Informationssystem genannt, welches informelles Wissen an Objekte der Wissensbasis koppeln kann. Zu Illustrationszwecken ist es auch möglich, an Diagnosen oder Symptomen Text–, Bild– oder HTML–Dokumente anzubinden. Diese können dann in der Dialogoberfläche von den Benutzern aufgerufen werden, um sich weitergehend über ein Thema zu informieren. In Abb. 3.8 sind einige Beispiele solcher Anbindungen gezeigt.

In diesem Abschnitt wurde der Expertensystembaukasten D3 mit den wichtigsten Funktionalitäten vorgestellt. Natürlich war es in diesem Rahmen nicht möglich, alle vorhandenen Komponenten und Konzepte vollständig vorzustellen oder alle Anwendungsalternativen zu zeigen. Für eine vollständige Beschreibung sei hier auf [PUPPE et al. 1996] verwiesen.

Im Folgenden wird nun die tutorielle Komponente von D3 beschrieben, die aus Wissensbasis und Fallsammlung die Generierung von fallbasierten Trainingssystemen ermöglicht. Wie in Abschnitt **??** beschrieben, müssen diese Systeme folgende Aktionen ermöglichen:

1. Präsentation von Falldaten.
2. Möglichkeit des Studenten, nach neuen Informationen zu fragen.
3. Interpretation dieser Merkmale, z. B. Erkennen von Merkmalen aus Bildern.
4. Stellen von Verdachtsdiagnosen auf Basis der bekannten Daten.

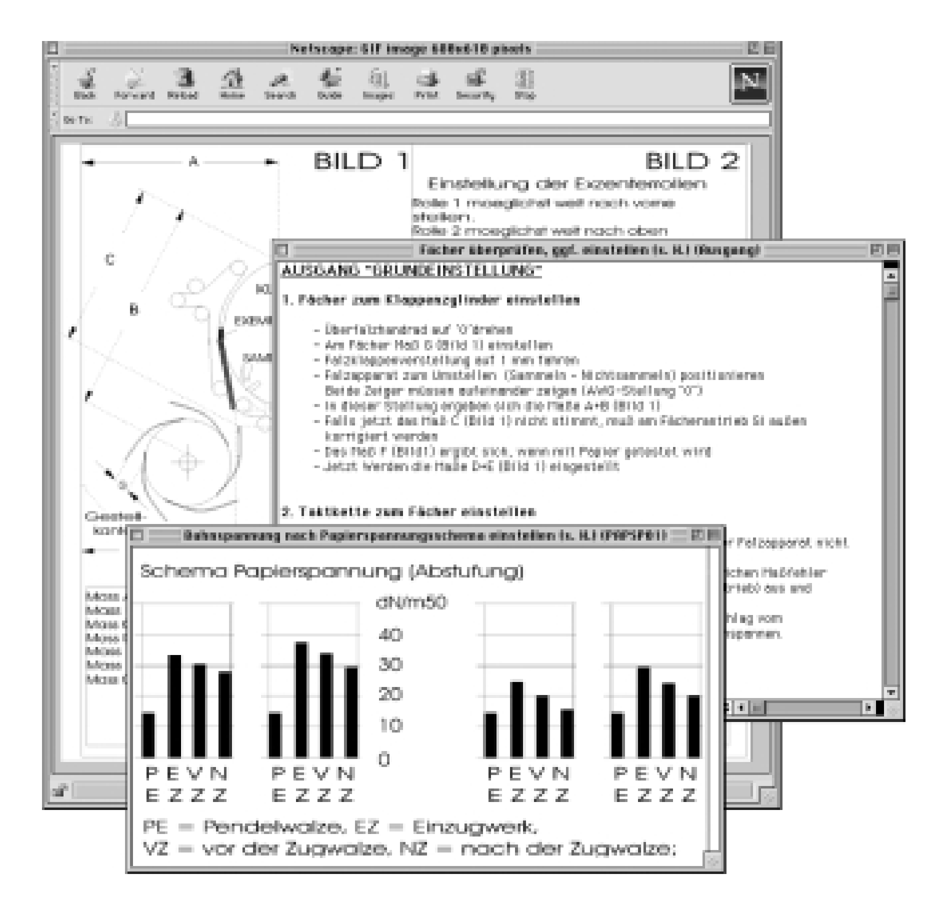


Abbildung 3.8: Das Beispiel aus der Fehlerdiagnostik von Druckmaschinen zeigt ein HTML–Dokument mit Schaltplänen zur Behebung bei einem Fehlerfall. In dem Textdokument sind die korrekten Schalterstellungen beschrieben, während in einem Bild die Papierspannungen aufgetragen sind. Mit diesen Hilfen kann ein Monteur vor Ort den Fehler besser beschreiben, finden und beheben, da er sofort genau auf die benötigten Informationen zugreifen kann, anstatt erst lange im Handbuch suchen zu müssen.

Die Aktionen werden in einem zeitlichen Wechselspiel notwendig, so müssen nach der Anforderung neuer Informationen (2) diese in der Fallpräsentation (1) erscheinen, wobei sie zur Bestätigung der Verdachtsdiagnosen (4) dienen und manche Merkmale erst aus einem Bild erkannt werden müssen (3). Alle möglichen Aktionen des Studenten sollten kritisierbar sein, wobei diese Kritik durch Erklärungen plausibel gemacht werden kann.

Zum besseren Verständnis des D3TRAINERS wird im Folgenden eine typische Fallbearbeitung durch einen Studenten in einer medizinischen Anwendung beschrieben.

3.2 Typischer Ablauf einer Fallbearbeitung

Beim Start eines Falles in einem Trainingssystem, das mit dem D3TRAINER generiert wurde, muß zuerst der zu bearbeitende Fall ausgewählt werden. Dies kann entweder nach Namen oder per Zufall erfolgen, sowie durch eine Auswahl nach bestimmten Kriterien, die ungefähr der allgemeinen Fallsuche in der Dialogumgebung des Expertensystems D3 entspricht.

Nach der Auswahl eines Falles werden dem Studenten die Grunddaten des Falles, bei dem Beispielfall die allgemeinen Patientendaten (*Alter*, *Geschlecht* usw.) in einer dynamischen Hierarchie präsentiert. Dieses Fenster (Abb. 3.9) bleibt für die ganze Fallbearbeitung der zentrale Mittelpunkt, in dem alle neuen Informationen eingetragen werden und aus dem heraus alle anderen Aktionen aufgerufen werden. Durch das selektive Auf– und Zuklappen von Frageklassen und Folgefragen kann der Student alle für ihn relevanten Symptome gleichzeitig sehen. Ein durchschnittlicher medizinischer Fall in der Rheumatologie besitzt schließlich etwa 80 Symptome, die unmöglich alle gleichzeitig angezeigt werden können. Für einen besseren Überblick über die Vielzahl an Symptomen sind diese in abnorm und normal eingeteilt und in der Fallpräsentation farbig unterschieden.

Gleichzeitig zum Präsentationsfenster erscheint ein Fenster, in dem einige informelle Daten des Falles beschrieben werden (siehe Abb. 3.10), aus denen auch Folgebilder oder –texte aufgerufen werden können. Diese informelle Beschreibung dient dazu, den Fall realistischer zu gestalten und den Patienten zusätzlich zu beschreiben.

Der Student kann sich nun die bisherigen Falldaten ansehen und sich weitere Informationen zu allen Symptomen aufrufen. Grundsätzlich ist es möglich, Erklärungen anzuzeigen, Symptome, wenn möglich, bildlich darzustellen, und begleitendes Lehrmaterial anzubieten (siehe Abb. 3.11). Der Student kann sich so ein besseres Bild von dem momentanen Stand des Patienten machen.

In der Standardumgebung kann der Student immer zwischen zwei mög-



Abbildung 3.9: In einer auf– und zuklappbaren Hierarchie werden alle Anfangssymptome (Hauptbeschwerden) dargestellt. Durch die Buttons *Information, Therapien*, *Diagnose* und *Abschluß* werden die anderen Aktionen aufgerufen.

lichen Aktionen wählen: eine neue Information anfordern oder Diagnosen verdächtigen. Die Reihenfolge wird nicht vorgegeben und ist alleine durch den Studenten bestimmt.

Bei der Entscheidung für neue Informationen wird dem Studenten eine Teilmenge möglicher Frageklassen (Abb. 3.12) angeboten, aus der er eine oder mehrere auswählen kann. Falls er sich nicht sicher ist, welche der Untersuchungen er als nächstes durchführen soll, kann er auch die *Systemwahl* abfragen, also die Frageklasse, die das darunterliegende Expertensystem als nächstes erfragen würde. Auf Basis dieser Frageklasse kann eine Bewertung der durch den Studenten ausgewählten Frageklasse stattfinden. Der Student begründet seine Auswahl an Frageklassen durch Angabe von Verdachtsdiagnosen, zu deren Klärung er diese Information anfordern möchte. Auch diese Begründung wird kritisiert, indem überprüft wird, ob die ausgewählten Diagnosen durch die Untersuchungen abgeklärt werden könnten und ob diese Diagnosen überhaupt verdächtig sind. Nach der Selektion von Frageklassen werden die Ergebnisse der gewählten Untersuchungen in dem Präsentationsfenster dargestellt und der Student kann sich so weiter über den Zustand des Patienten informieren. Zu jeder Untersuchung werden drei Größen (Kosten in DM, Patientenrisiko in Punkten und Zeitaufwand des Arztes in Minuten) verrechnet und in einer Aufwandsskala eingetragen (siehe Abb. 3.13). So erhält der Student einen Überblick über die momentanen Gesamtkosten des Falles, wobei er sich jederzeit die Detailauflistung ansehen kann.

Entscheidet sich der Student zur Selektion von Verdachtsdiagnosen, wer-



Abbildung 3.10: Informelle Darstellung der Patientin: In einer kurzen Beschreibung werden die Hauptbeschwerden zusammengefaßt. Durch Klick auf die Texte *Inspektion des Gesichts*, bzw. *Inspektion der Hände* werden eigene Bildfenster geöffnet, die genau diese Bereiche anzeigen.

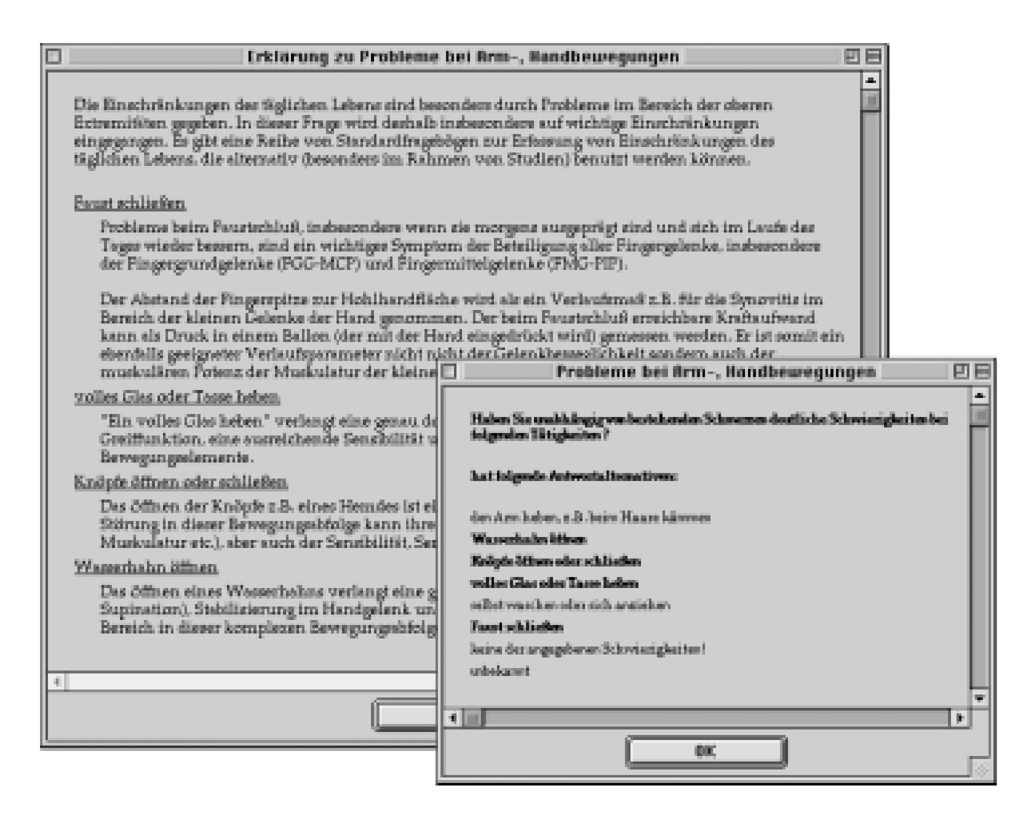


Abbildung 3.11: Erklärungsoptionen zu dem Symptom *Probleme bei Armbewegungen*: Hier werden in dem Fenster *Probleme bei Armbewegungen* der volle Fragetext und die möglichen Antwortalternativen angezeigt, die bei der Falleingabe angeboten wurden. In dem Fenster *Erklärung zu Probleme bei Armbewegungen* ist die Bedeutung dieser Frage und der vorliegenden Ausprägungen für den diagnostischen Prozeß beschrieben.

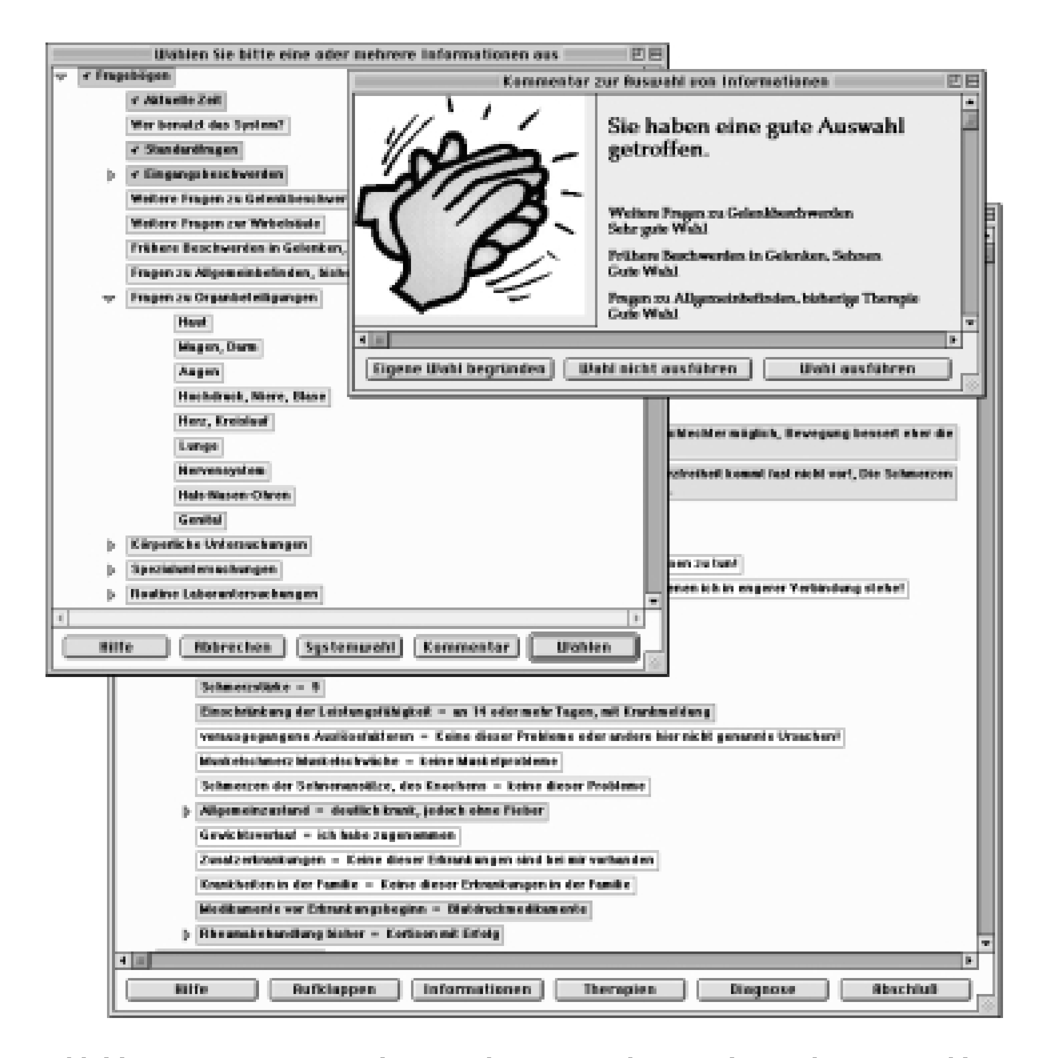


Abbildung 3.12: Aus einer dynamischen Hierarchie wurden mehrere Frageklassen selektiert und durch Vergleich mit der Wahl des Expertensystem bewertet. Anschließend werden die neuen Ergebnisse dieser Untersuchungen in das Präsentationsfenster (im Hintergrund) eingetragen.

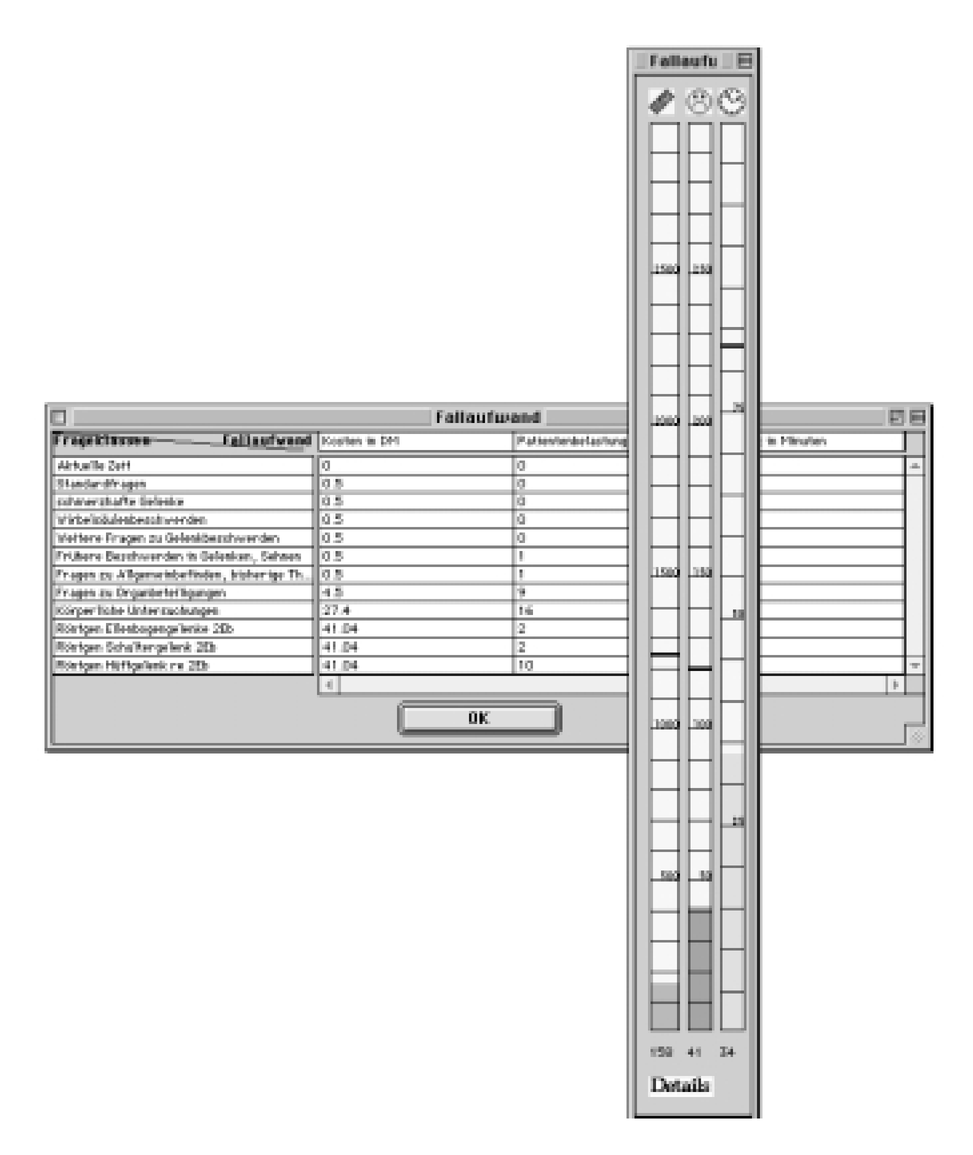


Abbildung 3.13: In einer Balkendarstellung werden die Gesamtsummen aller bisher entstanden Kosten dargestellt. Durch die Option *Details* werden in einer Tabelle die Gesamtkosten auf die angeforderten Untersuchungen verteilt, so daß der Student genau sehen kann, wieviel Kosten jede einzelne Information verursacht hat.

den ihm, ebenfalls in einer dynamischen Hierarchie, alle Diagnosen des Systems präsentiert (siehe Abb. 3.14). Hier kann er beliebig viele Diagnosen auswählen und sie in die Kategorien *verdächtig* und *bestätigt* einteilen. Auf der Basis der Systemdiagnosen werden nun die Benutzerdiagnosen verglichen und das Ergebnis dieser Bewertung wird in einem Feedback–Fenster dargestellt. Auch hier kann der Student verschiedene Erklärungsfunktionen aufrufen, um sich diese Bewertung begründen zu lassen (mehr dazu in Abschnitt 3.6).

Es können neben Diagnosen auch Therapien zur Behandlung dieser Diagnosen ausgewählt werden. Bei der typischen Fallbesprechung ist eine getrennte Auswahl der beiden Typen vorgesehen, wobei der Ablauf bei der Therapiebehandlung grundsätzlich analog zu der Behandlung von Diagnosen verläuft.

Als Sonderform von Symptomen gibt es Bildinterpretationen, die von dem Studenten erkannt werden müssen. Diese Symptome sind in der Fallpräsentation farblich gekennzeichnet. Möchte der Student eines dieser Symptome interpretieren, so ruft er die Bildinterpretation auf (Abb. 3.15). Hier kann er alle Bilder, die das Symptom beschreiben, in einer Karteikastenstruktur betrachten, während auf der linken Seite die Möglichkeit gegeben wird, erkannte Symptomatiken einzugeben. Der Student kann sich die korrekte Lösung für die Bildinterpretation ansehen, auf Basis derer er kritisiert wird. Nach Abschluß seiner Bildinterpretation werden die von ihm erkannten Symptome in den Fall übernommen und erscheinen in dem Präsentationsfenster.

Dieses Wechselspiel von Informationsselektion, Stellen von Verdachtsdiagnosen und Bildinterpretation kann der Student über die verschiedenen Phasen einer Fallbearbeitung (im medizinischen Beispiel Anamnese, körperliche Untersuchung, Labor und technische Untersuchungen) durchführen, bis er den Fall gelöst hat. Zum Fallabschluß muß er seine finalen Diagnosen und Therapien angeben, die zusammen mit dem Fallaufwand und der Bildinterpretation in einer Gesamtbewertung (Abb. 3.16) kritisiert werden. Im Gegensatz zu den bisherigen Bewertungen werden die Benutzeraktionen nun mit der echten Fallösung verglichen, und nicht wie bisher nur auf Basis der momentanen Fallsituation.

Nach der Abschlußbewertung mit ihren unterschiedlichen Erklärungsoptionen erscheint eine Fallbesprechung (siehe Abb. 3.17), d. h. ein Text, in dem der Experte seine abschließende Einschätzung des Falles beschreibt. Nach Beenden des Falles kann eine Fallverlaufsbewertung, ein Protokoll aller durch den Studenten aufgerufenen Kritiken und Hilfeaufrufen, in eine Datei gesichert werden.

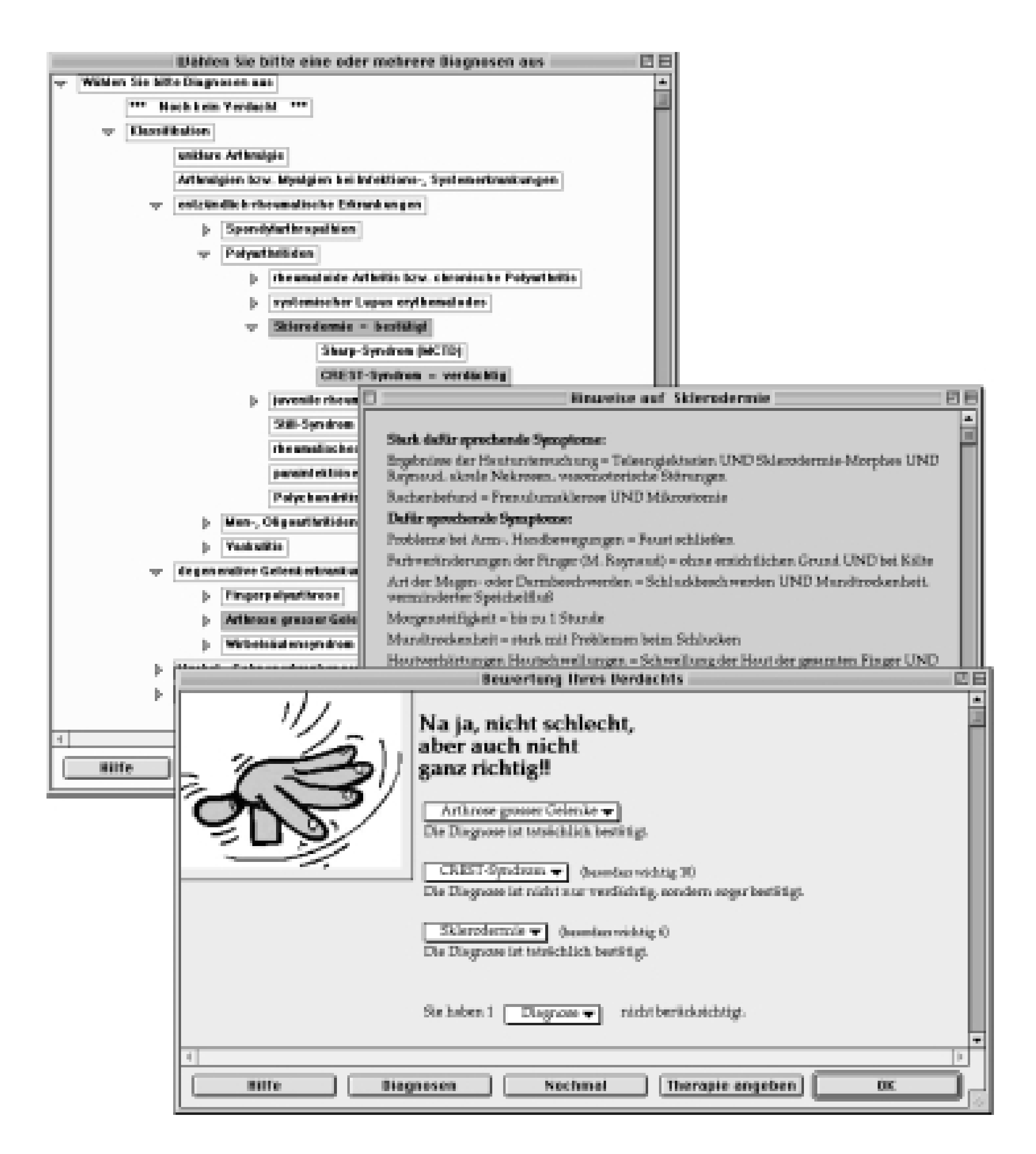


Abbildung 3.14: Aus einer dynamischen Hierarchie aller Diagnosen wählt der Student bestätigte und / oder verdächtige Diagnosen aus, die dann in einem Feedback–Fenster (unten) bewertet werden. Dazu wird eine grobe Einteilung der Gesamtleistung mit einem Bild symbolisiert und die Bewertung der Einzeldiagnosen angezeigt. Zu jeder Diagnose können nun die hinweisenden Symptome des Falles und alle grundsätzlich auf diese Diagnose zeigenden Symptome angezeigt werden.

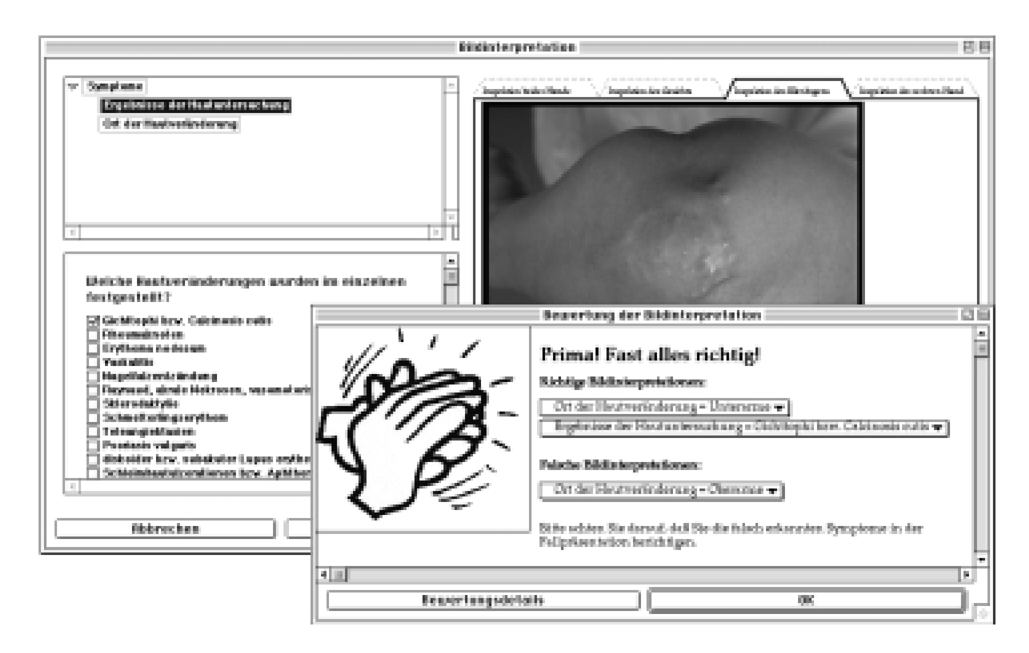


Abbildung 3.15: In dem Bildinterpretationsfenster sind rechts die verschiedenen Bilder zur Beschreibung des Symptoms in einem Karteikasten gesammelt. Zu jedem Bild erscheint links oben das zu erkennende Symptom, welches links unten durch den Studenten eingegeben werden kann. Durch den Button *Lösung* kann er sich die Lösung zu dem aktuellen Bild ansehen, während durch den Button *Kritik* seine eingegebene Interpretation bewertet wird. Nach Schließen der Bildinterpretation durch *OK* werden die von ihm erkannten Symptome in die Falldarstellung übernommen.

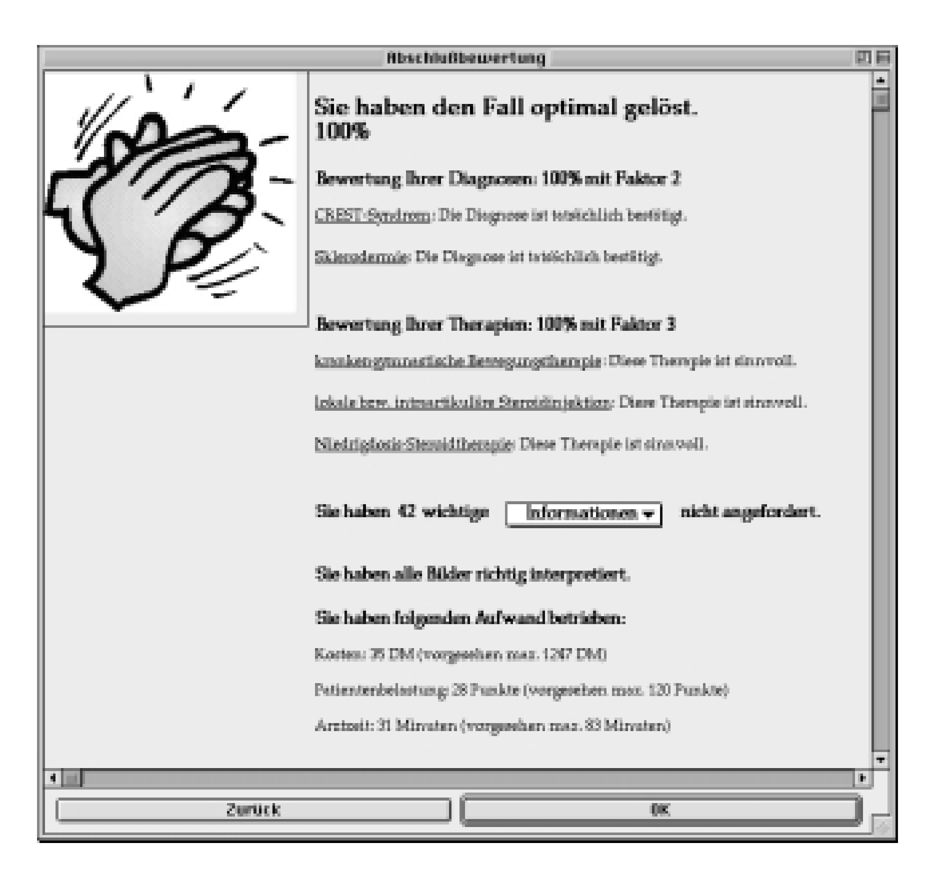


Abbildung 3.16: Zum Abschluß des Falles erhält der Student eine Bewertung seiner Gesamtlösung (Diagnosen, Therapien, Frageklassenauswahl und Bildinterpretation) mit entsprechenden Erklärungsoptionen. Basis dieser Bewertung ist nicht wie sonst die aktuelle Fallsituation sondern der Gesamtfall.

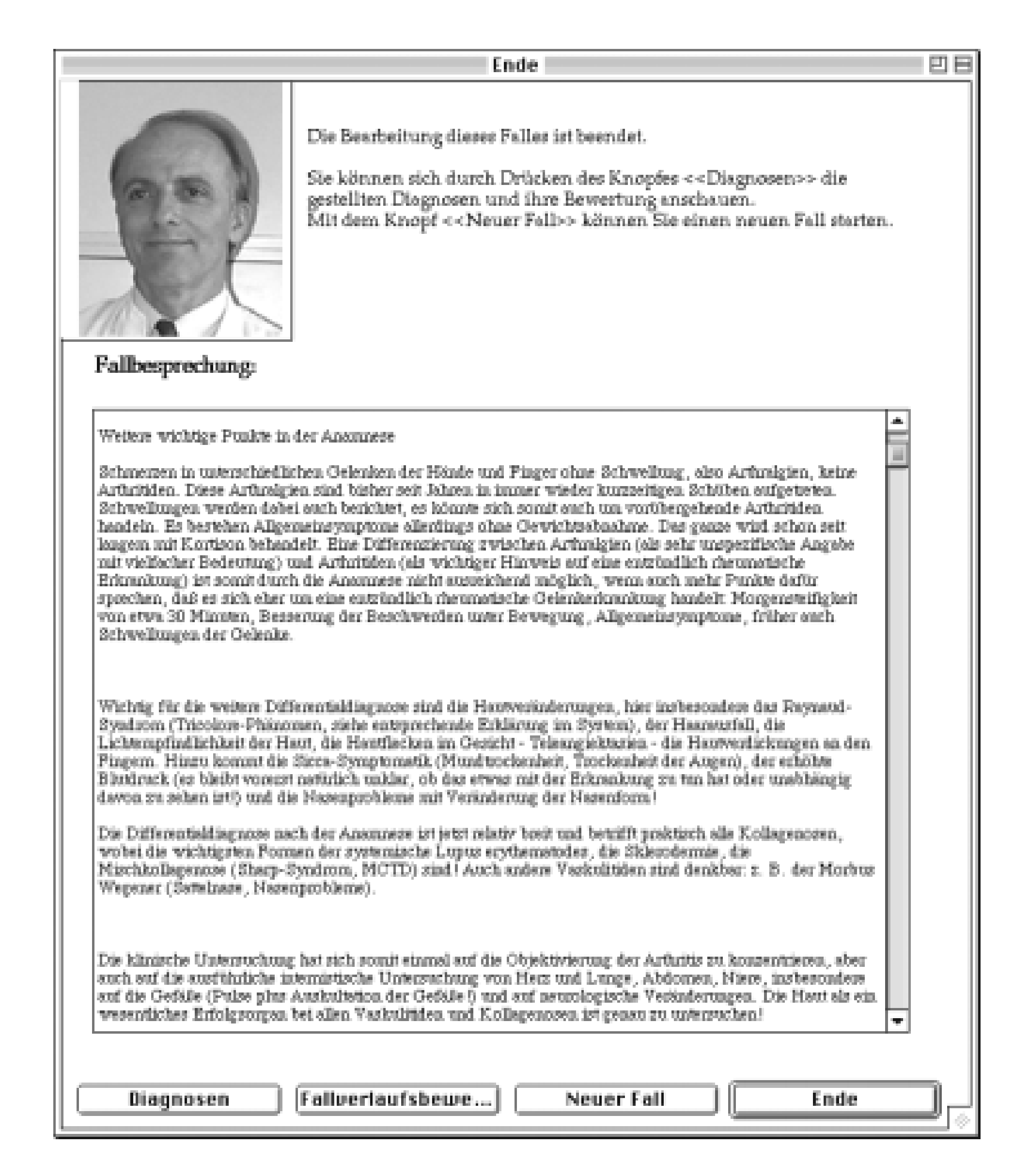


Abbildung 3.17: Im Anschluß an die Bearbeitung des Falles wird dem Studenten ein Kommentar des Experten (Bild links oben) angezeigt, in dem dieser den Fall nochmals abschließend bespricht und auf Besonderheiten aufmerksam macht. Durch die Buttons kann eine Fallverlaufsbewertung angezeigt werden, wie auch die finalen Diagnosen des Falles nochmals genannt und erklärt werden.

3.3 Aufbau des Wissensmodells

Nachdem nun ein erster Eindruck entstanden ist, was in Trainingssystemen des D3TRAINERS möglich ist, wird jetzt darauf eingegangen, welche technischen Voraussetzungen vorliegen müssen, damit ein didaktisch sinnvolles System entstehen kann. Im Folgenden werden die drei Grundelemente *Wissensbasis*, *Fallbasis* und *Bildsystem* betrachtet, die man für die Generierung benötigt. Dabei wird gezeigt, welche Teile schon durch die Konsultation vorgegeben sind und welche Wisseneingaben speziell für die tutorielle Nutzung gemacht werden sollten. Grundsätzlich soll hier aber nochmals betont werden, daß der D3TRAINER aus jeder Konsultationswissensbasis und einer dazugehörigen Fallbasis ein Trainingssystem generieren kann, ohne daß die Notwendigkeit einer zusätzlichen Wisseneingabe besteht.

3.3.1 Wissensbasis

Eine Wissensbasis in D3, die für eine Anwendung in der Konsultation erstellt wurde und nun für die tutorielle Anwendung genutzt werden soll, muß bestimmte Eigenschaften haben und sollte um einige Attribute erweitert werden, um für einen *optimalen* Einsatz in der Lehre geeignet zu sein.

Die in den Trainingssystemen vorwiegend verwendete Wissensart sind heuristische, mehrstufige Regeln mit den Operatoren *und*, *oder*, *n–aus–m* und *nicht*. Da im D3TRAINER diese Regeln auf bloße Symptome–Diagnosen–Beziehungen reduziert werden, erscheint es in einigen komplexen Domänen sinnvoll, die Erklärung, warum eine Diagnose durch die vorhandenen Symptome erklärt wird, aus dem Wissen des überdeckenden Problemlösers zu generieren. Überdeckende Regeln haben im Gegensatz zu der ‚Symptomkombination–zeigt–auf–Lösung'–Struktur der heuristischen Regeln eine ‚Lösung–wird–erklärt–durch–Symptom'–Struktur, die dann bei der Erklärung sinnvoller eingesetzt werden kann. Dies verlangt natürlich, daß die beiden Problemlöser gut aufeinander abgestimmt sind, d. h. sie müssen zu den selben Diagnoseergebnissen kommen, da sonst eine heuristisch hergeleitete Diagnose nicht überdeckend erklärt werden kann. Grundsätzlich ist es aber auch möglich, den gesamten Vergleich auf den überdeckenden Problemlöser umzustellen.

Besitzt die Wissensbasis auch allgemeines Wissen, wie Abnormitätsaussagen oder Gewichtungen von Symptomen und Frageklassen, so kann dies nicht nur für das Auffinden ähnlicher Fälle für den Lehrbetrieb nützlich sein. Diese Attribute werden ebenfalls dazu genutzt, die Präsentation des Falles für den Studenten effektiver zu gestalten. Abnormitätsaussagen teilen die vorhandenen Symptome mit ihren Ausprägungen in normal und abnorm ein, so daß bei der Falldarstellung die Falldaten in unterschiedlicher Farbe dargestellt werden können oder die normalen Daten vollständig ausgeblendet werden. Betrachtet man die gesamte Masse an Informationen, die bei einem komplexen Fall dargestellt werden, so kann eine solche



Abbildung 3.18: Definition von Aufwandsgruppen durch die Aufwandsklassen *Anamnese* (Aufwandsklasse 1), *Körperliche Untersuchung* (Aufwandsklasse 2), *Labor* (Aufwandsklasse 3) und *Technische Untersuchung* (Aufwandsklasse 4) einer medizinischen Wissensbasis.

Hilfestellung entscheidend für die erfolgreiche Fallösung durch den Studenten sein.

Wissensbasen, die bisher nur in der Konsultation eingesetzt wurden, besitzen oft Fragen, die zur reinen Steuerung des Dialoges dienen und keine inhaltliche Funktion haben. Erscheinen diese Frageklassen (manchmal auch Symptome) für die Darstellung des Falles für den Studenten als sinnlos, dann kann der Wissensbasisentwickler das Attribut *Gewicht* des fallbasierten Problemlösers so einstellen, daß die Frageklasse in der Fallpräsentation des Trainingssystems ausgefiltert wird.

Eine Variante in den fallbasierten Trainingssystemen des D3TRAINERS ist die Präsentation der Falldaten in Gruppen. In einer medizinischen Umgebung wären das beispielsweise die Anamnese, körperlichen Untersuchungen, Laborbefunde und schließlich die technischen Untersuchungen. Eine schon in der Expertensystemshell eingebaute Funktionalität ist die Einstufung der Frageklassen in sogenannte Aufwandsklassen. Die Aufwandsklasse ist ein numerischer Wert in beliebiger Höhe, der bei der Dialogführung berücksichtigt wird, so daß immer Frageklassen niedriger Aufwandsklassen bei der Auswahl der nächsten zu stellenden Frageklasse bevorzugt werden. Für die Trainingsumgebungen können nun verschiedene Aufwandsklassen zu einer Aufwandsgruppe zusammengefügt werden (siehe Abb. 3.18), die dann auch zur optische Gliederung der Frageklassen benutzt werden.

Eine weitere Zusatzinformation, die bei den Frageklassen angesiedelt wird, ist die Eingabe von Kosten, Risiko und Zeitaufwand durch den Experten. Er kann den drei Größen eigene Namen und Einheiten geben und bestimmen, wie viele davon eingetragen werden (siehe Abb. 3.19). So können in einer Wissensbasis alle drei Attribute bei der Kostenbetrachtung der Testauswahl betrachtet werden, wohingegen bei einer anderen Wissensbasis



Abbildung 3.19: Einstellungen der Wissensbasis für den D3TRAINER: Die Trainer–Startfrageklassen sind die Frageklassen, die bei einem Fallstart präsentiert werden, ohne daß der Student sie anfordern muß. Darunter können Verbalisierung und Einheiten zu den drei Kostengruppen *Kosten*,*Zeitaufwand* und*Risiko* eingeben werden.

nur der Zeitfaktor berücksichtigt wird. Der Experte gibt die Werte zu den einzelnen Frageklassen in einer Attributtabelle ein. Diese Kostenattribute werden dann bei der Fallbearbeitung durch den Studenten aufsummiert und stellen eine weitere Bewertungsmöglichkeit dar.

Die Abstraktion von Rohdaten (Symptomen) zu Merkmalsabstraktionen (Symptominterpretationen) stellt einen Weg dar, den diagnostischen Mittelbau zu formulieren. Eine weitere Möglichkeit besteht in der Einführung sogenannter Kontextdiagnosen. Diese werden im diagnostischen Prozeß wie normale Diagnosen behandelt, erscheinen aber nie in der Ergebnisausgabe. Einige Kontextdiagnosen, die für die Konsultation eingeführt wurden, haben einen tutoriellen Wert, d. h. es erscheint sinnvoll, daß sie auch dem Studenten für seinen Problemlöseprozeß zur Verfügung stehen, wohingegen andere Kontextdiagnosen eine rein technische Begründung haben. Um in dem Trainingssystemen eine klare Unterscheidung zu finden, wurde eine weitere Klasse an Diagnosen eingeführt, die Tutorkontextdiagnosen. Sie werden in der Konsultation wie eine normale Kontextdiagnosen behandelt und erscheinen in dem Trainingssystem nicht. Die bisherigen Kontextdiagnosen, die keine Tutorkontextdiagnosen sind, werden in der Trainingsumgebung wie normale Diagnosen behandelt.

Besonderer Augenmerk muß auf die Eingabe von Erklärungen und Erläuterungen zu den einzelnen Diagnosen und Symptomen gerichtet wer-

den, da diese nicht mehr nur Hilfe bei der Benutzung des Systems sind, sondern in einem Trainingssystem zu Lerninhalten werden. Viele Erläuterungen, die zur Konsultation eingeben wurden, sind ohne Änderungen auch in der Lernumgebung sinnvoll. Bei manchen Erläuterungen kommt es aber vor, daß sie entweder sinnlos erscheinen oder dem Lernprozeß sogar abträglich sein können. Beispielsweise können hier Erklärungen zum Umgang mit dem Eingabeformat der Frage gegeben werden. In diesen Fällen kann in dem Wissenserwerb eine zweite Ebene der Erklärung eingetragen werden, die sogenannte Tutorerklärung. In der Ablaufumgebung wird nun zu jedem Objekt überprüft, ob eine solche Tutorerklärung existiert. Wenn keine Tutorerklärung angegeben ist, wird die ‚normale' Erklärung benutzt. Grundsätzlich müssen alle erklärenden Anbindungen, wie auch die Bilder, Texte und HTML–Dokumente, daraufhin überprüft werden, ob sie gewissen pädagogischen Kriterien genügen.

Die in diesem Abschnitt genannten Erweiterungen müssen nicht durchgeführt werden. Sie führen jedoch zu einem besseren, didaktisch sinnvolleren System mit erweiterten Möglichkeiten beim Lernen. Der zusätzliche Aufwand für den Bereichsexperten zur Eingabe dieser Informationen ist bei einer existierenden Wissensbasis gering und verändert die Wissensbasis in anderen Anwendungsszenarien (z. B. der Konsultation) nicht.

3.3.2 Fallbasis

Die Expertensystemshell D3 besitzt verschiedene Dialogformen, mit denen Fälle und deren Lösungen abgespeichert werden können. Bei dem Abspeichervorgang werden neben Symptomen, deren Ausprägung und den durch das System hergeleiteten Diagnosen auch andere Attribute des Falles abgespeichert. Dies sind organisatorische Daten wie beispielsweise Autor, Datum und Zeit, aber auch inhaltliche Informationen, wie die Einschätzung des Autors, ob die Systemdiagnosen korrekt hergeleitet wurden.

Abb. 3.20 zeigt ein Fallformular, welches beim Abspeichern eines Falles erscheint und in dem der Experte das Fallzusatzwissen eingeben muß. Neben der Einteilung der Systemdiagnosen in verschiedenen Kategorien (z. B. wichtige Hauptdiagnose, fragliche Nebendiagnose) kann er jede Diagnose mit einem numerischen Faktor versehen, der die Wichtigkeit der Diagnose in diesem konkreten Fall widerspiegelt. Alle Diagnosen ohne Faktor haben die Wichtigkeit eins. Diese Faktoren werden im D3TRAINER bei der Bewertung der Benutzerdiagnosen verrechnet, so daß der Student nicht stark kritisiert wird, wenn er eine Nebendiagnose nicht angegeben hat.

Neben dem Kurzkommentar, der in wenigen Zeilen in dem Fallformular eingegebenen werden kann, besteht die Möglichkeit, daß der Experte eine ausführliche Fallbesprechung zu dem aktuellen Fall bereitstellen möchte. In dieser Fallbesprechung werden dann die Besonderheiten und Schwie-



Abbildung 3.20: Formular zum Abspeichern eines Falles. Neben *Fallname, Autor* und *Datum* können noch *Schwierigkeitsgrad* und *Fallart* eingegeben werden, die später für die Fallsuche interessant werden. *Systemdiagnosen korrekt?* und *Tatsächliche Diagnosen* sind Aussagen über die zutreffenden Lösung des Falles. Hier werden auch die Einstufung der korrekten Diagnosen und ihre Wichtigkeit in diesem Fall (Faktor) festgelegt. Die notwendigen Frageklassen des Falles werden in dem D3TRAINER als Basis für die Bewertung der unbedingt durchzuführenden Frageklassen benutzt, das *angebundene Bildsystem* (vgl. Abschnitt 3.3.3) ist ein Dateiverweis, ebenso wie die *Fallbesprechung*. Der hier angegebene *Kommentar* ist ausschließlich für den Experten gedacht und wird den Studenten im D3TRAINER nicht gezeigt.

rigkeiten des Falles aus der Sicht des Experten diskutiert. Die Fallbesprechung ist eine eigene HTML–Datei, auf die nur der Verweis gespeichert wird.

Ebenfalls nur der Verweis auf eine Datei ist die Angabe eines Bildsystems, welches für die multimediale Präsentation eines Falles in der Trainerumgebung benötigt wird. Wie diese Bildsysteme aufgebaut werden, ist in Abschnitt 3.3.3 beschrieben.

Diese mehr technische Betrachtung der Falleingabe darf nicht darüber hinwegtäuschen, daß die Auswahl und Bearbeitung der Fälle eine der wichtigsten Aufgaben der Experten bei der Generierung der Trainingssysteme ist. Die bereitgestellten Fälle stehen im Mittelpunkt des Lernprozesses und müssen von dem Expertensystem einwandfrei bearbeitet werden können, da Unstimmigkeiten im dynamischen Wissen des Systems dem Studenten ebenfalls vermittelt werden.

3.3.3 Bildsysteme

Ein Bildsystem ist für die multimediale Präsentation des Falles zuständig, wie sie später in dem Trainingssystem erfolgt (siehe auch [REINHARDT und REMP 1995], [REINHARDT 1996]). Der erste Schritt ist sicherlich die Sammlung von Bild– und Textelementen, die Symptome und ihre Ausprägungen in diesem speziellen Fall gut beschreiben. In dem generierten Trainingssystem muß der Lernende schließlich diese Symptomausprägungen aus den Bildern erkennen.

Ein wichtiger Bestandteil eines Bildsystems ist das sogenannte Startbild (siehe Abb. 3.21), das einen ersten Eindruck von dem Fall vermitteln soll. In medizinischen Anwendungen ist das eine Kurzbeschreibung des Patienten, bei der Pflanzenwissensbasis die Totalansicht der zu klassifizierenden Pflanze. In dem Startbild werden bestimmte Regionen markiert, an die Nachfolgebilder angebunden werden können. So kann man zum Beispiel eine Detailansicht, Vergrößerungen oder einen Schnitt durch eine Stelle anbieten.

Das Erstellen von Bildsystemen erfolgt durch graphische Editoren, so daß die Bereichsexperten auch diese Aufgabe ohne Programmierkenntnisse erfüllen können.

Ein Bild des Bildsystems hat folgende Eigenschaften (siehe Abb. 3.22):

- einen Namen,
- relative Pfadangabe zur Bild–Datei,
- (optional) die Angabe, welche Symptomausprägungen man darauf erkennen kann und
- (optional) eine Menge an ausgezeichneten Regionen mit folgenden Eigenschaften



Abbildung 3.21: Das Startbild eines Bildsystems dient zur Vermittlung des ersten Eindrucks eines Falles. Hier ist es eine kurze Beschreibung der Hauptsymptome. An die klicksensitiven Stellen *Inspektion des Gesichts* und *Inspektion beider Hände*, die mit den Werkzeugen der Palette (linker Fensterrand) eingetragen wurden, werden über Pop–Ups weitere Bilder angehängt.

- (optional) die Angabe, welche Symptomausprägungen man hier erkennen kann,
- (optional) Nachfolgebilder.

Neben Bildern können auch Texte (siehe Abb. 3.23) eingeben werden, die den Fall besser erklären. In den Texten können ebenfalls sensitive Textbereiche definiert werden, an die Symptomausprägungen oder Folgebilder angebunden werden.

Ein Bildsystem besteht also aus einer Menge von Bild– und Textelementen mit einem ausgezeichneten Startbild. Die Elemente sind partiell miteinander über Nachfolgerrelationen verbunden und die Elemente und Regionen in den Bildern und Texten können mit Symptomausprägungen verknüpft werden, die dort erkannt werden können. Zum Überblick über ein bestehendes Bildsystem kann ein Navigationsgraph geöffnet werden (siehe Abb. 3.24), in dem nochmals alle Verknüpfungen aufgezeigt werden.

Die Erstellung solcher Bildsysteme benötigt meist etwas Zeit und Sorgfalt, da ja die Abstimmung mit dem Fall berücksichtigt werden muß. Aus diesem Grund wurde eine Generierung dieser Bildsysteme aus schon vorliegendem Bildmaterial und einem neuem Fall umgesetzt. Da die Bilder miteinander verträglich sein müssen, wurde eine zusätzliche Beschreibung der Bildelemente eingeführt, die z. B. sicherstellt, daß bei einer weiblichen Patientin kein männlicher Schultergürtel ausgewählt wird. Natürlich kann die Generierung nur dann sinnvoll sein, wenn das Wissensgebiet es erlaubt. Beispielsweise ist es in der Medizin nicht wichtig, ob das Röntgenbild nun wirklich von demselben Patienten stammt, solange es dieselben Veränderungen an denselben Gelenken hat. In der Pflanzenwissensbasis ist aber zum Beispiel eine Generierung nur schwer möglich, da die Bilder ja genau eine Pflanze, also genau diesen Fall darstellen und nicht durch Bilder anderer Pflanzen ersetzbar sind.

Nach der vollständigen Bearbeitung eines Bildsystems wird dieses in eine Datei gesichert und in dem Formular zum Abspeichern eines Falles eingebunden, so daß es bei der Generierung eines Trainingssystem mit Verwendung dieses Falles eingebunden werden kann.

3.4 Interaktivität

Der Grad der Interaktion in einem fallbasierten Trainingssystem wird festgelegt durch die Art und Anzahl der Möglichkeiten des Lernenden, in das Geschehen, also die Fallbearbeitung, einzugreifen. Ein hoher Grad wird erreicht, wenn alle Teilaufgaben

- *Symptomsuche* (Auswahl von Untersuchungen),

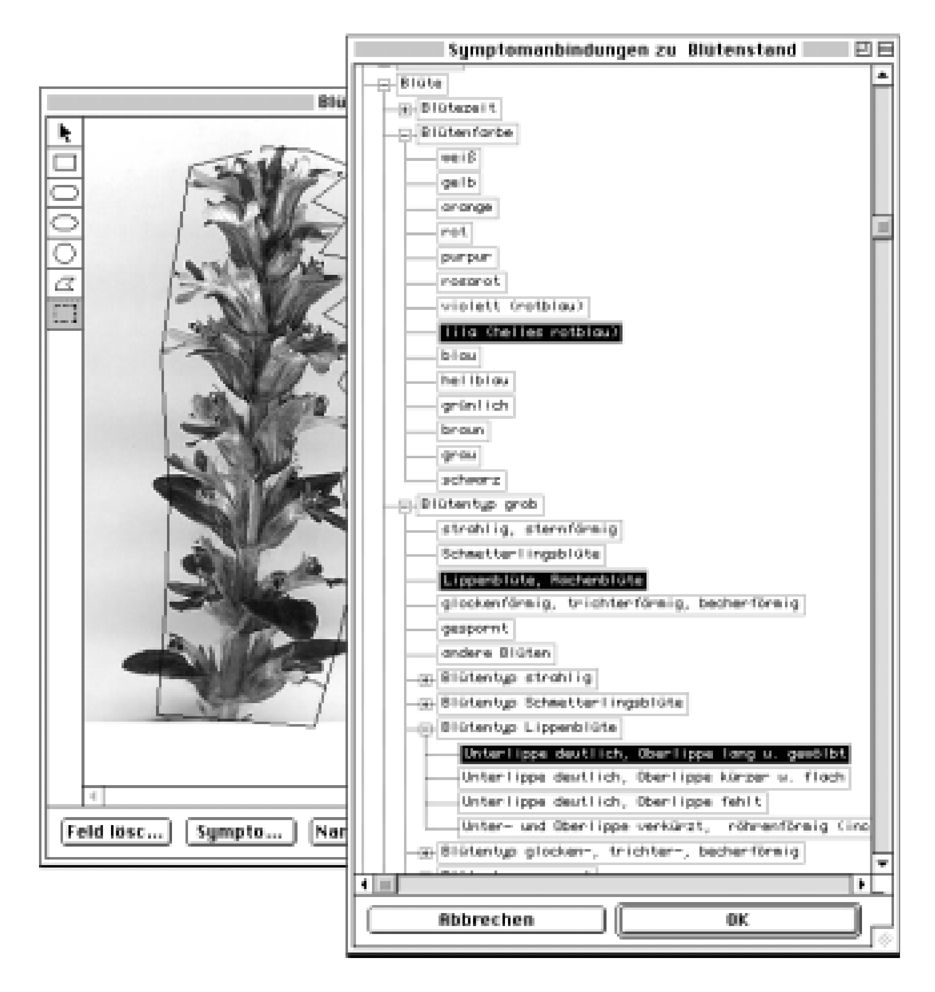


Abbildung 3.22: Innerhalb eines Bildsystems können zu einzelnen Bilder wieder Regionen definiert werden, an die weitere Bilder, Texte oder Symptome angehängt werden. Das Anhängen von Symptomen geschieht über eine aufklappbare Hierarchie, aus denen die Symptome ausgewählt werden, die in diesem Bild erkannt werden können.

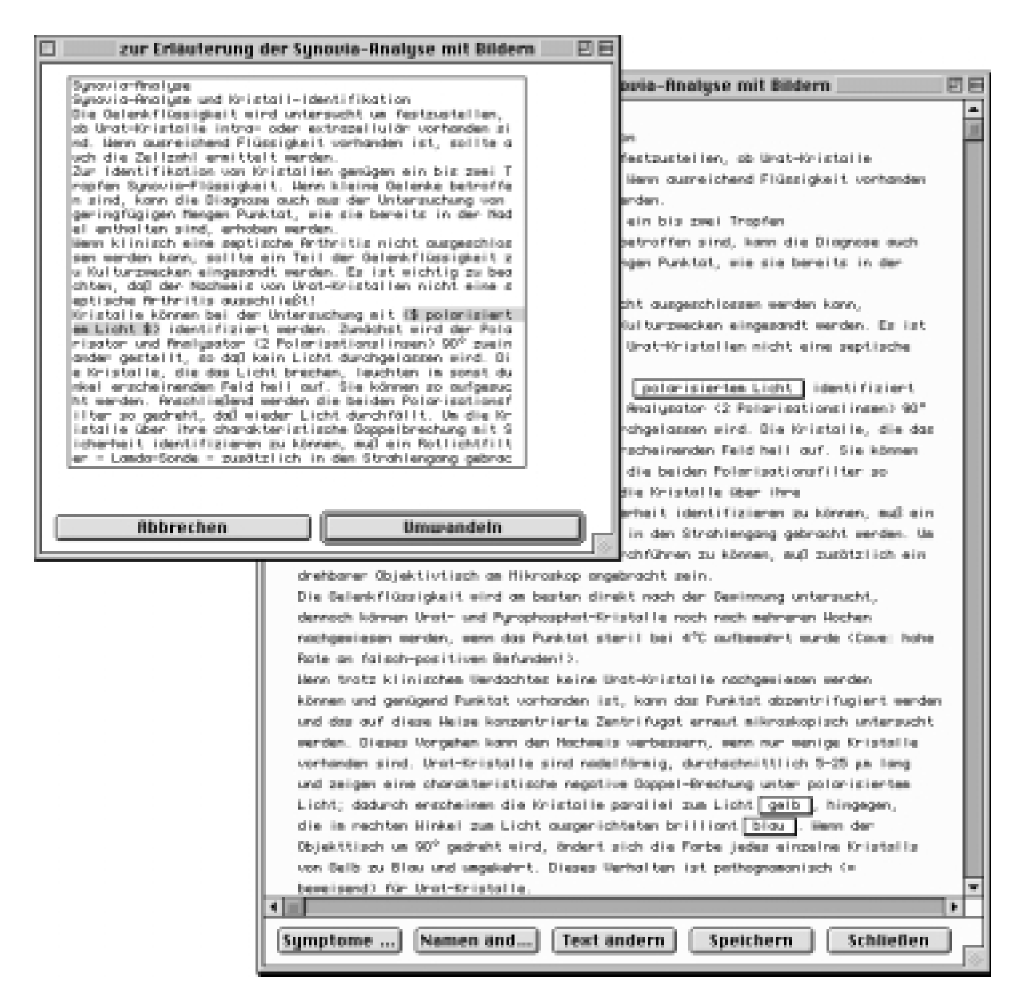


Abbildung 3.23: In einem Bildsystem können auch reine Texte abgelegt werden, die ebenfalls klicksensitive Wortbereiche haben. In einem Eingabefenster (links) wird der Text und darin die klicksensitiven Bereiche (durch Klammerung mit ($ und $)) editiert. In dem Textfenster selbst können an diese Bereiche wiederum andere Bilder oder Symptome angebunden werden.

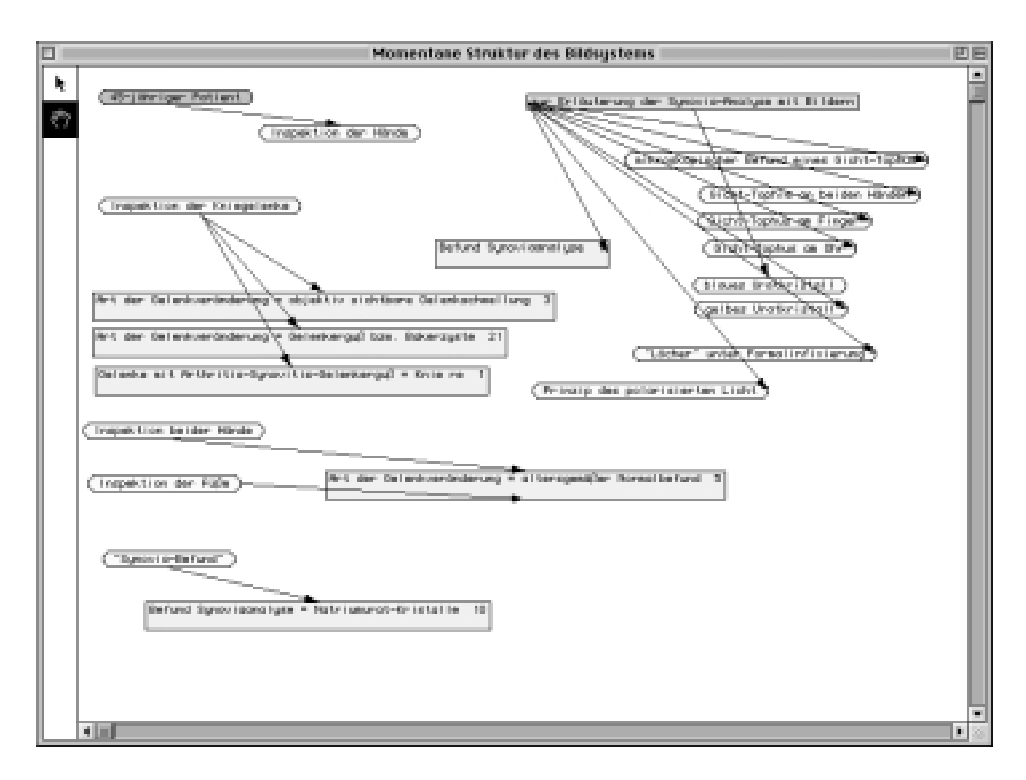


Abbildung 3.24: Zur Übersicht über ein Bildsystem wird dem Autor ein Strukturgraph angeboten. Die Rundecke sind Bilder, die Rechtecke Texte oder Symptome (farblich unterschieden). Pfeile zwischen Bildern und Texten bedeuten, daß hier ein Aufruf von einem Bild aus dem anderen erfolgen kann, während Pfeile von Bildern auf Symptome anzeigen, daß diese Symptome in dem Bild erkannt werden können. Die Zahlen hinter den Symptomen sind Gewichtungen, die bei der Kritik der Bildinterpretation des Studenten verwendet wird. Sie sind ein Hinweis darauf, wie wichtig das Merkmal ist und wie gut man es erkennen kann.

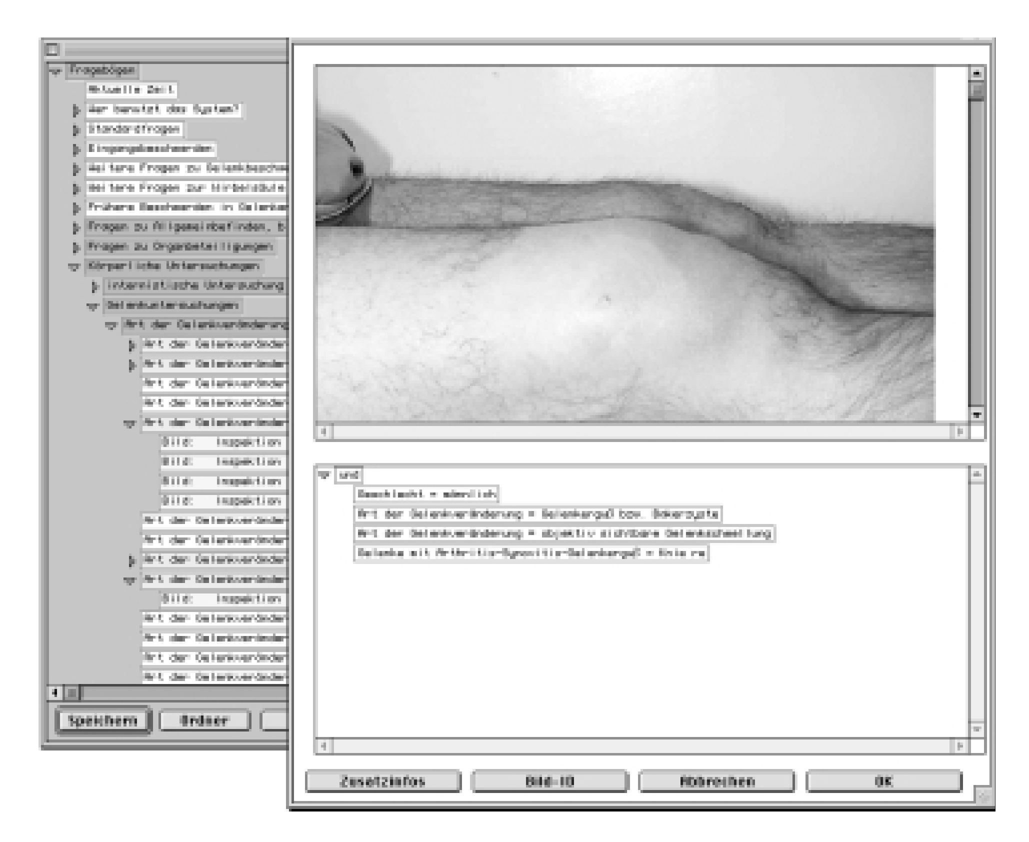


Abbildung 3.25: Aus einer Menge bereits eingegebener Bildsysteme kann eine Generierungshierarchie (im Hintergrund) gewonnen werden. Sie zeigt alle Symptome der Wissensbasis an und welche Bilder die Symptomausprägungen darstellen können. Durch Klick auf ein Bild öffnet sich ein Detailbild (rechts), in dem noch zusätzliche Attribute eingegeben werden können, die für die Generierung wichtig sind (z. B. das Geschlecht des Patienten). Man kann den Bildern auch eine Bild–ID geben und Bilder des gleichen Patienten mit derselben ID kennzeichnen. Bei einer Generierung werden dann Bilder mit der gleichen Bild–ID bevorzugt oder ausschließlich verwendet.

- *Symptomerhebung* (Erkennen von Merkmalen) und
- *Symptominterpretation* (Schließen aus Fakten auf Lösungen)

durch den Benutzer durchgeführt und begründet werden können. Der zweite Aspekt bei der Bestimmung der Interaktivität ist die Auswahl der Reihenfolge der Aktionen. Legt der Benutzer die Reihenfolge selbst fest, so kann von einer höheren Interaktivität gesprochen werden, als bei einem System, in dem das System die Abarbeitungsfolge festlegt. Bei diesen Überlegungen darf aber nicht angenommen werden, daß ein hoher Interaktionsgrad immer besser für die momentane Lernumgebung ist. In manchen Situationen ist der didaktisch günstigere Weg, den Lernenden durch die Fallbearbeitung zu leiten und seine Interaktivität möglichst gering zu halten.

Die Entscheidung für den sinnvollen Interaktionsgrad hängt also stark von den Lernvoraussetzungen ab, in denen das System eingesetzt werden soll. Aus diesen Grund bietet der D3TRAINER mehrere Ebenen an Interaktionsmöglichkeiten an, die entweder durch den Fachbereichsexperten oder den Lernenden selbst eingestellt werden. Die erste Entscheidung ist hier, ob der Lernende durch den Fall geführt wird oder die Abarbeitungsreihenfolge selbst bestimmt. Die Einschränkung der Aufgaben, die der Lernende bei der Fallösung durchführen soll, ist eine weitere Möglichkeit zur Steuerung der Interaktivität. Hier können von einfacher Hypothesenbildung für die Lösungen bis zu Symptomerkennung und Untersuchungsanforderung verschiedene Kombinationen gewählt werden, die den Schwerpunkt bestimmen. Auch die Strenge der Bewertung fließt in das Design der Fallbearbeitung ein: Es ist möglich, daß der Lernende zwar mehrere komplexe Aufgaben bewältigen muß, die Kritik seiner Aktionen jedoch sehr zurückhaltend formuliert wird, um ihn nicht durch sehr strenge Kritik zu demotivieren.

Ein weiterer Bereich, der bei der Beurteilung der Interaktionsfähigkeit eines Systems betrachtet werden muß, ist der Zeitpunkt, an dem der Lernende durch das System kritisiert wird. Bezüglich des **Bewertungszeitpunktes** beim D3TRAINER können zwei verschiedene Varianten gewählt werden: Der Übungs– und der Prüfungsmodus. In dem Prüfungsmodus werden grundsätzlich keine Bewertungen während der Fallbearbeitung gegeben, d. h. der Student bekommt nur am Ende seiner Fallösung eine zusammenfassende Kritik. In dem Übungsmodus wird noch zwischen zwei Aktionen unterschieden: Aktionen, denen eine unmittelbare Kritik folgt und Aktionen, die nur auf Anforderung durch den Studenten kritisiert werden. Unmittelbare Kritik folgt auf Selektion von Verdachtsdiagnosen und Therapien sowie auf die Begründungen dieser Selektion durch Angabe von Symptomen und Diagnosen. Bei der Auswahl von neuen Informationen, bei deren Begründung durch Angabe von Verdachtsdiagnosen und auch bei der

Eingabe von Bildinterpretationen (Symptomerkennung aus Bildern) kritisiert das Trainingssystem den Studenten erst nach dessen Aufforderung. Diese Unterscheidung wurde gemacht, um die wichtigste diagnostische Aufgabe, die Diagnosenfindung, in den Mittelpunkt zu stellen, während die anderen Aufgaben der Diagnostik (Symptomerkennung und Nachfragen) nicht ständig überprüft werden. Diese Zweiteilung verkürzt die Bearbeitungszeit eines Falles erheblich und hilft so, die Motivation des Studenten zu erhalten.

Die Umsetzung von speziellen didaktischen, stark geführten Dialogformen in dem D3TRAINER wurde in einer Diplomarbeit experimentell untersucht [BERGSTEINER 1995]. Beispielhaft wurde versucht, den sokratischen Dialog auf die Abarbeitung des D3TRAINERS umzusetzen. Die Grundidee des sokratischen Dialoges, implizites Wissen durch Fragen und mittels Beispielen ohne Vermittlung von Wahrheiten explizit zu machen, wurde in einer Reihe von Regeln kodiert. Beispielhaft wurde eine Wissensbasis zum Reisanbau, dem Standardbeispiel aus Why [WENGER 1987], angelegt und Beispielsitzungen mit dieser und einer komplexen medizinischen Wissensbasis durchgeführt. Während sich der Dialog bei dem Reisanbau tatsächlich nach dem sokratischem Grundgedanken verhielt, stellte sich bald heraus, daß diese Dialogform für die medizinische Anwendung nicht praktikabel war. Gründe dafür waren die komplexen Fälle mit ca. 80 Symptomen, die in ihrer Gesamtheit nur teilweise im Gedächtnis behalten wurden und bei denen man auch nur schwer Vergleichsfälle zu dem Problem des aktuellen Falles fand. Bei beiden Wissensbasen mußte der Dialog aber auch oft durch die Angabe der richtigen Lösung abgebrochen werden, wenn das System in einer Situation endete, in der keine Fragen mehr zu stellen waren, da keine Vergleichsfälle zur Verfügung standen. Dieser Abbruch durch Vermittlung von Wissen steht allerdings im Widerspruch zu der sokratischen Grundidee. Trotzdem zeigte dieses Experiment, daß es grundsätzlich möglich ist, eigene Dialogformen in die Bearbeitung eines Falles in dem D3TRAINER einzubauen.

Als letzte Bemerkung zu der Interaktivität sei hier die Benutzungsoberfläche aufgeführt, die ja die Kommunikation und somit auch die Interaktivitätsmöglichkeiten des Systems bestimmt. Die Benutzungsoberfläche des D3TRAINERS ist rein mausgesteuert, d. h. der Student benötigt zur Fallbearbeitung keine Tastatureingabe, sondern kann alle Aktionen durch Klick auf entsprechende Buttons anstoßen. In dem gesamten Trainingsprogramm wurde hoher Wert auf die Einheitlichkeit der Interaktionsmöglichkeiten gelegt. So gibt es beispielsweise als durchgängiges Bedienelement die auf– und zuklappbaren Hierarchien, die für den Studenten immer die gleiche Bedienung ermöglichen. Die Oberfläche kann generell adaptiert werden, indem alle verwendete Farben und Bilder (z. B. Feedbackbilder, Bilder für Kosten, Aufwand und Risiko) frei gewählt werden.

Es besteht auch die Möglichkeit, eine komplett eigene Benutzungsoberfläche zu programmieren, die über Schnittstellen mit den anderen Komponenten des D3TRAINERS kommuniziert. Ein Beispiel für eine solche Implementation des Java–Clients des D3TRAINERS [FAULHABER und REINHARDT 1997]. Durch eine Client–Server–Architektur [FAULHABER 1997] wird die gesamte Benutzeroberfläche auf einen WWW–Browser umgelenkt, der über CGI–Interaktion mit dem Server Daten austauscht. Denkbar ist, über den selben Mechanismus die interne Funktionalität des D3TRAINERS zu umzulenken und eine eigene Oberfläche, z. B. stark abgestimmt auf das Anwendungsgebiet, zu benutzen.

Die gesamte Oberfläche und somit auch die Interaktionsmöglichkeiten könnten auch für eine Anwendung vollständig gekapselt werden, wie es beispielsweise für die Bereitstellung im WWW umgesetzt wurde. Somit kann, allerdings mit Programmieraufwand, auch eine neue Oberfläche für eine spezielle Domäne entstehen.

Für das Design der Interaktivität in einem System gibt es meist keine eindeutig richtige Entscheidung für alle Anwendungsszenarien. Deswegen ist es auch wichtig, möglichst variabel auf die unterschiedlichen Bedürfnisse eingehen zu können. Der D3TRAINER bietet hierfür eine Vielzahl an Optionen bereits an, die normalerweise genügend Freiraum für die verschiedenen Anforderungen der Lernumgebungen lassen. In speziellen Lernumgebungen kann es allerdings vorkommen, daß eine eigene Benutzeroberfläche notwendig ist. In diesem Fall kann die Funktionalität des D3TRAINERS über definierte Schnittstellen angesprochen werden.

3.5 Präsentationskomponente

Bei der Präsentation des Falles für den Studenten stehen im D3TRAINER alle vier in Abschnitt **??** genannten Variationen zur Verfügung, d. h. durch Setzen der Einstellungen können folgende Modi zur **Symptomanforderung** benutzt werden:

- Der **ganze Fall** wird dem Studenten zu Beginn vollständig präsentiert. Er muß keine Symptomanforderung durchführen und kann sich völlig auf die restlichen diagnostischen Aufgaben *Schlußfolgern* und *Symptomerkennung* konzentrieren.
- Alle Fragegruppen des Falles sind in Aufwandsgruppen eingeteilt, die dem Studenten nacheinander präsentiert werden. Nach jeder neuen Fragegruppe muß er einen Diagnoseverdacht äußern, bevor er zu der nächsten Gruppe kommen kann. In diesem Modus muß er die Symptomanforderung noch nicht selbst durchführen, wird aber in die **schrittweise Bearbeitung des Falles** eingeführt. Die Anzahl der

Aufwandsgruppen kann durch den Bereichsexperten frei festgelegt und die Gruppen selbst benannt werden.

- In diesem Modus wird die Symptomanforderung vollständig **lernergesteuert** durchgeführt. Der Student kann zu jedem Zeitpunkt selbst bestimmen, welche Fragegruppen er als nächstes stellen möchte und kann sich diese Auswahl auch kritisieren lassen. Hier wird auch der Zeitpunkt der anderen Aktionen, z. B. Diagnoseverdacht stellen, nur durch den Studenten bestimmt.

- Eine Erweiterung des letzten Modus stellt die **Einzelanforderung** von Fragen einer Fragegruppe dar. Hier wird dem Studenten nicht die gesamte Frageklasse präsentiert, sondern er bekommt nur alle Fragestellungen dieser Fragegruppe angezeigt und kann sich dann aus diesen die gewünschten Einzelfragen heraussuchen. Dieser Modus hat den Vorteil, daß der Student spielerisch herausfinden kann, wie wichtig die einzelnen Fragen für die gesamte Problemlösung sein können. Natürlich sind nicht alle Fragegruppen für diese Einzelanforderung geeignet, so daß sie in der Wisseneingabe durch ein eigenes Attribut gekennzeichnet werden.

In den beiden zuletzt genannten Modi werden auch die entstehenden **Kosten** der Fragegruppenauswahl aufsummiert, so daß der Student eine Übersicht über den Aufwand seiner Problemlösung bekommen kann. Es können bis zu drei Kostenarten eingegeben werden, die in Einheit und Namen frei durch den Wissensbasisentwickler bestimmt werden. In medizinischen Anwendungen sind das beispielsweise Kosten in DM, Risiko für den Patienten in Punkten und aufgewendete Arztzeit in Minuten.

Sowohl bei der schrittweisen Anforderung der Fragegruppen, wie auch bei der Einzelanforderung, wird dem Studenten eine **Startinformation** gezeigt, auf die er aufbauend weitere Fragegruppen auswählen soll. Diese Startinformationen können in dem D3TRAINER die in der Konsultation festgelegten Startfrageklassen sein, es können aber auch Fragegruppen speziell für die tutorielle Anwendung angegeben werden.

Bei der **Symptompräsentation** muß generell unterschieden werden, ob zu einem Fall ein Bildsystem (wie in Abschnitt 3.3.3 beschrieben) angegeben wurde. Existiert ein solches, können alle Symptome, die in den medialen Elementen des Bildsystems erkennbar sind, durch den Studenten eingegeben und somit kritisiert werden. Ist zusätzlich zu den einzelnen Symptomen Abnormitätswissen eingegeben, welches in jeder Wissensbasis mit fallvergleichender Problemlösung benötigt wird, so kann der D3TRAINER auch noch zwischen normalen und abnormen Symptomausprägungen unterscheiden. In diesem Fall können alle in Abschnitt **??** genannten Optionen

zur Symptompräsentation angeboten werden. Im einzelnen sind das Kombinationen aus den folgenden didaktischen Entscheidungen:

Normale und abnorme Symptome?

- Normale Symptome werden ausgefiltert, d. h. sie werden dem Studenten überhaupt nicht angezeigt.
- Normale Symptome werden dem Studenten zwar angezeigt, sind aber farblich von den abnormen Symptomen unterschieden.
- Normale und abnorme Symptome sind für den Studenten farblich nicht unterschieden.

Bildinterpretation?

- Die medialen Elemente des Bildsystems stellen nur eine Ergänzung der Symptompräsentation dar und müssen durch den Studenten nicht erkannt werden.
- Die medialen Elemente des Bildsystems müssen durch den Studenten interpretiert werden und seine Interpretation, also die Symptomerkennung, wird durch das System bewertet.

Aus diesen Entscheidungen können wie in Abschnitt **??** gesagt alle genannten Variationen der Symptompräsentation kombiniert werden, so daß verschiedene Schwierigkeitsstufen bzw. Lernbedingungen für den Studenten berücksichtigt werden können.

Bei der **Symptomverbalisierung** wird im D3TRAINER die einfache *Symptom–Wert*–Verbalisierung benützt, da sie vor allem bei großen Falldaten übersichtlicher als vollständige Sätze ist. Eine Umsetzung auf natürlichsprachlichen Text durch Schablonen ist zwar technisch umsetzbar, wurde aber bisher nicht geplant. Für umfangreiche multiple–choice Fragen mit vielen Antwortalternativen stehen noch zwei zusätzliche Funktionalitäten bereit, die der Übersichtlichkeit dienen. Erstens gibt es in der Erklärung zu einem Symptom den Punkt *Information zur Frage*, der den vollständigen Fragetext bei der Falleingabe sowie alle möglichen Antwortalternativen aufzeigt, wobei alle Alternativen, die in diesem konkreten Fall zutreffen, durch Fettdruck markiert sind. Zweitens gibt es wenn möglich die Option, daß die Antwort zu einer Frage durch ein Bild dargestellt wird. Sinnvoll erscheint das beispielsweise bei der Angabe von betroffenen Gelenken, die dann nicht als Auflistung der Gelenkbezeichnungen, sondern als markierte Gelenke auf einem Skelett dargestellt werden (siehe Abb. 3.26). Diese Umsetzung ist aber natürlich nicht bei allen Symptomen möglich.

Zusammenfassend kann gesagt werden, daß der D3TRAINER aus allen verfügbaren Wissensbasen eine variable Symptompräsentation generieren

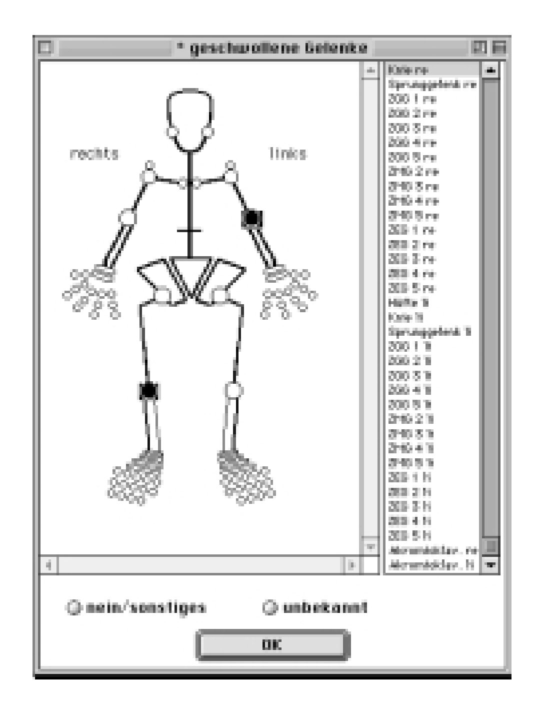


Abbildung 3.26: Bei manchen multiple–choice Merkmalen ist es möglich, die Werte nicht nur verbal, sondern, wie hier bei dem Gelenkmännchen gezeigt, auch über ein Bild darzustellen. Diese Visualisierung der Antwortalternativen trägt zur Übersichtlichkeit bei und erspart dem Studenten Zeit bei der Erfassung der Fallproblematik.

kann. Um aber zu einem optimalen Ergebnis zu gelangen, müssen folgende Zusatzinformationen in der Wissensbasis angegeben sein:

- Eine Unterscheidung in normale und abnorme Symptome ist nur dann möglich, wenn die entsprechenden Attribute zu jedem Symptom eingegeben wurden. Dieses Wissen kann schon in der Konsultationswissensbasis eingegeben worden sein, da es ja auch für die fallvergleichende Problemlösung wichtig ist.

- Bei der schrittweisen Einführung des Studenten in die Fallbearbeitung, müssen die Aufwandsgruppen zu dieser Domäne festgelegt werden.

- Soll eine Symptomerkennung durchgeführt werden, muß zu jedem Fall ein Bildsystem erstellt werden. Dies geschieht entweder durch den Bereichsexperten oder mittels Bildsystemgenerators.

- Stimmen die Startfrageklassen der Konsultation nicht mit den tutoriell sinnvollen Startfrageklassen überein, so müssen diese in der Wissensbasis zusätzlich angegeben werden.

- Für eine Berücksichtigung der benötigten Kosten, Zeit und Risiken müssen diese Größen in der Wissensbasis deklariert, sowie die entsprechenden Werte bei jeder Fragegruppe eingetragen werden.

Die Visualisierung von multiple–choice Antwortalternativen durch ein Bild ist durch die grundsätzliche Funktionalität der Expertensystemshell D3 gegeben, die es erlaubt, im Dialog Fragen durch Selektieren von ausgezeichneten Regionen in einem Bild zu beantworten. Diese Bilder und die entsprechenden Regionen können dann in dem entstehendem Trainingssystem ohne Zusatzaufwand benutzt werden.

Im Gegensatz zu den gerade vorgestellten Erklärungen sind die statischen Erklärungen unabhängig von dem aktuellen Fall.

In dem D3Trainer sind einige Erklärungsoptionen bereitgestellt, um dem Studenten das statische Wissen der Wissensbasis näherzubringen. Sie stehen dem Lernenden, ähnlich wie ein Lehrbuch, zur Verfügung, um inhaltliche Informationen zu gewinnen, die er für die Fallbearbeitung nutzen kann. Es werden folgende Optionen angeboten:

Fragetext und Antwortalternativen zu Symptomen: Der Student kann sich anschauen, wie die Frage in dem Konsultationsdialog gestellt wurde und welche Antwortalternativen dem Fallautor zur Verfügung standen (siehe Abb. 3.11). Oft kann dies bei Terminologieproblemen helfen, wenn der Student zwischen verschiedenen Belegungen des Symptoms nur schwer unterscheiden kann.

Kurzerläuterungen und Normbereiche zu Symptomen: Die Kurzerläuterungen sind kurze Texte, die kleine Probleme bei dem Umgang mit dem Symptom verhindern können. Beispielsweise kann das eine Anmerkungen sein, ob das dargestellte Bild spiegelverkehrt ist. Bei numerischen Symptomen können hier zusätzlich die Normalbereiche dargestellt werden.

Spezielle Kurzerklärungen zu den aktuellen Werten zu Symptomen: Zu jedem Symptom und seinen Antwortalternativen können Erläuterungen eingegeben werden, die die diagnostische Bedeutung der Belegung klarstellt (siehe Abb. 3.11).

Angebundene externe Bilder, Texte und HTML–Dokumente: Diese Erklärungen decken den größten Teil des Lehrbuchwissens der Trainingssysteme ab (siehe Abb. 3.8). Es können nicht nur HTML–Dokumente, sondern auch ganz allgemein Links zu WWW–Anwendungen angegeben werden, so daß sogar Verknüpfungen zu kompletten Lehrbüchern im Internet bereitgestellt werden können.

Der D3TRAINER ist also fähig, alle in Abschnitt **??** genannten didaktischen Varianten über Einstellung durch den Fachexperten oder gegebenenfalls durch die Studenten selbst umzusetzen. Die einzige Ausnahme bleibt dabei die Transformation der *Symptom–Wert*–Verbalisierung auf natürlichsprachliche Konstrukte, was aber einen erhöhten Aufwand bei der Wissensbasisentwicklung bedeutet und nicht zur Übersichtlichkeit der Fallpräsentation beiträgt. Die Umsetzung der Fallpräsentation in dem D3TRAINER durch dynamisch auf- und zuklappbare Hierarchien wurde gerade deshalb gewählt, um die enorme Menge an Symptomen pro Fall in eine für den Studenten angenehme Umgebung zu bringen. Die Einführung von natürlichsprachlichen Sätzen würde diesen Vorteil wieder zunichte machen.

3.6 Die Kritikkomponente

Die Kritikkomponente des D3TRAINERS ist eine sehr variabel gestaltbare und mächtige Umgebung, die durch Einstellungen im Trainingssystem, durch Variablen der Wissensbasis und durch Vorgaben in den einzelnen Fällen gesteuert wird (siehe auch [REINHARDT und PUPPE 1997]). Diese Attribute werden teilweise durch den Bereichsexperten vorgegeben, können aber zum Teil auch durch den Studenten bestimmt werden. Der D3TRAINER stellt dem Studenten in einem Trainingssystem folgende kritisierbare Aktionen zur Verfügung, die anschliessend im Detail besprochen werden:

- Selektion von Verdachtsdiagnosen

- Selektion von Therapien
- Begründung der Verdachtsdiagnosen durch Auswahl von hinweisenden Symptomen und Diagnosen
- Selektion neu anzufordernder Informationen
- Begründung der angeforderten Informationen durch Angabe von Verdachtsdiagnosen, die durch diese Informationen geklärt werden sollen
- Erkennen von Merkmalen aus Bildern

In allen Bereichen der Kritikkomponente wird das Konzept verfolgt, daß die intern berechneten Punkte dem Studenten nicht direkt vermittelt werden. Die Punktzahlen werden auf bis zu sechs Kategorien abgebildet, die dem Studenten durch einheitliche Bewertungsbilder und –texte sehr schnell vermitteln, wie gut die durchgeführte Aktion war. Dieses Konzept hat den Vorteil, daß der Student sich nicht lange mit der detaillierten Berechnung seiner Bewertung beschäftigt, sondern sich auf das Wesentliche, seine Lösung, konzentriert. Die im Folgenden aufgestellten Algorithmen entstanden in Zusammenarbeit mit den Fachexperten und wurden in verschiedenen Iterationen verbessert. Die angebenen Schätzwerte haben sich in den verschiedenen Stadien und Anwendungen als günstig erwiesen und wurden an vorgegebenen Datenstrukturen des Expertensystembaukastens D3 angelehnt.

3.6.1 Selektion von Verdachtsdiagnosen

In einem mit dem D3TRAINER generierten Trainingssystem kann bzw. muß der Student zu verschiedenen Zeitpunkten eine Diagnosenauswahl treffen, die dann durch das System mittels eines komplexen Vergleichs mit der Lösung des Expertensystems bewertet wird.

Der erste wichtige Aspekt bei dem Design des Systems ist die Entscheidung, wie der Student die **Auswahl** an Diagnosen bzw. Therapien treffen kann. Der D3TRAINER bietet dafür zwei Optionen an: Aus den korrekten Diagnosen und einer zufälligen Auswahl anderer Diagnosen wird eine multiple–choice Auswahl (siehe Abb. 3.27) generiert, aus der der Student die passenden Objekte selektiert. Natürlich stellt das eine enorme Erleichterung der Aufgabe dar und kann deshalb nicht als Standardeinstellung benutzt werden. Die zweite Option ist die Auswahl von Objekten aus der Menge aller Diagnosen, die der Übersichtlichkeit wegen in einer geordneten Hierarchie dargestellt werden, die der Student durch Auf– und Zuklappen durchsuchen kann (siehe Abb. 3.14). In beiden Fällen stehen den

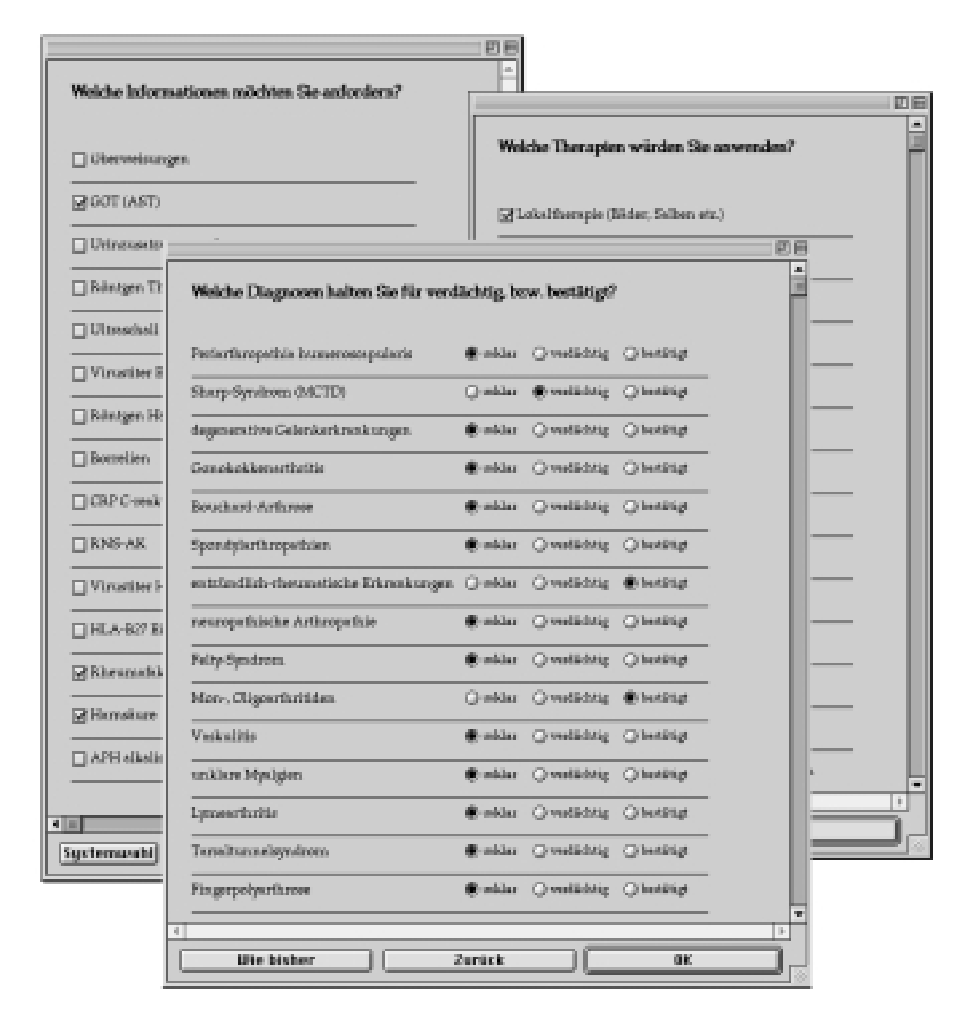


Abbildung 3.27: Auswahl von Untersuchungen, Diagnosen und Therapien aus einer Multiple–choice Auswahl.

Studenten zwei Kategorien, *verdächtig* und *bestätigt*, als **Sicherheitsfaktoren** zur Verfügung, mit denen er beschreiben kann, wie sicher er sich ist, daß diese Diagnosen in dieser Fallsituation zutreffen.

Da die Systemlösung des D3–Expertensystems auch in Kategorien geordnet ist, wäre die naive Lösung ein direkter Abgleich der beiden Mengen Systemdiagnosen und Benutzerdiagnosen. Dies würde aber einige Probleme mit sich bringen, die auch die Bewertung didaktisch in Frage stellen würden. Die meisten wurden schon in Abschnitt **??** genannt, so daß sie auch im D3TRAINER berücksichtigt werden sollten.

Ein wichtiger Punkt bei der Bewertung der ausgewählten Diagnosen ist, ob **benachbarte oder ähnliche Diagnosen** zu der Systemlösung selektiert wurden. Der D3TRAINER kann hier Ober– und Folgediagnosen erkennen,

Systemdiagnose	Studentendiagnose	Punkte
Systemdiagnose bestätigt	Diagnose bestätigt	100 %
Systemdiagnose bestätigt	Diagnose verdächtigt	75 %
Systemdiagnose bestätigt	Oberdiagnose bestätigt	50 %
Systemdiagnose bestätigt	Oberdiagnose verdächtigt	25 %
Systemdiagnose bestätigt	Folgediagnose bestätigt	50 %
Systemdiagnose bestätigt	Folgediagnose verdächtigt	50 %
Systemdiagnose verdächtigt	Diagnose bestätigt	75 %
Systemdiagnose verdächtigt	Diagnose verdächtigt	100 %
Systemdiagnose verdächtigt	Oberdiagnose bestätigt	75 %
Systemdiagnose verdächtigt	Oberdiagnose verdächtigt	25 %
Systemdiagnose verdächtigt	Folgediagnose bestätigt	25 %
Systemdiagnose verdächtigt	Folgediagnose verdächtigt	25 %

Tabelle 3.1: Prozentzahlen für die Abbildung der Punktzahl auf benachbarte und ähnliche Diagnosen.

für die dann ein Teil der erreichbaren Punktebewertung anerkannt wird. So wird es dem Studenten mit einer bestimmten Prozentzahl der erreichbaren Punkte gewertet, wenn dieser statt der verdächtigen Systemdiagnose die Oberdiagnose als bestätigt angibt. Die Einteilung in Ober– bzw. Folgediagnosen ist in allen Wissensbasen explizit durch die Ordnung in einer Hierarchie gegeben, die der Fachexperte bei dem Aufbau der Wissensbasis von Beginn an benötigt. Somit entsteht kein zusätzlicher Aufwand, um eine solche Ähnlichkeit zweier Diagnosen zu errechnen. Genau sind die Ähnlichkeitswerte in Tabelle 3.1 angegeben. Die dadurch errechneten Punktzahlen werden dann in dem Bewertungsalgorithmus berücksichtigt.

Der erwähnte Bewertungsalgorithmus ist besonders wichtig, da Trainingssysteme, die mit dem D3Trainer erstellt wurden, fähig sind, mit **Mehrfachdiagnosen** umzugehen. Sowohl das System als auch der Student können zu jeder Zeit mehrere Diagnosen als Fallösung angegeben. Dies ist besonders wichtig, da in der Realität selten medizinische Fälle auftreten, deren Symptomatik auf nur einer Ursache (Diagnose) beruht. Wenn man reale Fälle in eine problemorientierte Lernumgebung einbinden will, muß man aber vor allem darauf achten, daß das Lernziel klar erkennbar ist. Bei einer Fallbesprechung muß also zwischen Neben– und Hauptdiagnosen des Falles unterschieden werden, um die Aufmerksamkeit des Studenten auf den gewünschten Bereich der Problemlösung zu lenken. Der D3Trainer bietet für dieses Problem an, zu jedem Fall die Lösungen durch Kategorien oder Faktoren zu gewichten (siehe Abb. 3.20), so daß diese bei dem Bewertungsalgorithmus unterschiedlich gewichtet werden können. Grundsätzlich reicht eine Einteilung in die Kategorien *Hauptdiagnose*, *Nebendiagnose*

und *Nebendiagnose fraglich* aus, um die Fokussierung des Studenten entsprechend zu steuern. Wird eine differenziertere Unterscheidung benötigt, so kann der Fallautor die Diagnosen mit ganzen Zahlen beliebiger Höhe gewichten, die dann miteinander verrechnet werden. Diese Gewichtungen sind nur fallabhängig und haben keinen Einfluß auf andere Fälle oder die Wissensbasis, auch sind sie für die anderen Anwendungsbereiche des Expertensystems ohne Bedeutung.

Die aktuelle **Fallsituation** ist ein weiterer Punkt bei der Bewertung der Verdachtsdiagnosen des Studenten. Grundsätzlich kann unterschieden werden, ob die Aktionen des Lernenden mit denen des Systems verglichen werden, die dieses in der aktuellen Fallsituation vorschlägt oder denen, die in der Gesamtheit des Falles richtig sind. Der Vorteil des *allwissenden* Trainingssystem ist, daß berücksichtigt werden kann, wenn der Lernende aufgrund nicht ausgewählter Untersuchungen oder falscher Symptomerkennung zwar richtig gefolgert hat, aber die korrekte Falldiagnose damit nicht finden kann. Optional kann hier weiter auf den fehlerhaften Daten geschlossen werden (aktuelle Fallsituation) oder der Lernende wird durch das zusätzliche Wissen des Systems auf die fehlerhaften oder fehlende Eingaben aufmerksam gemacht. Der D3TRAINER bietet beide Möglichkeiten durch eine Einstellung an, so daß sich entweder Wissensbasisautor oder Studenten für eine didaktische Variante entscheiden können.

Durch die starke Integration der einzelnen Teilaufgaben der Diagnostik (Stellen von Verdachtsdiagnosen, Testanforderung, Symptomerkennung) kommt es manchmal zu Problemen, wenn beispielsweise eine falsch erkannte Bildinterpretation auf eine andere als die korrekte Verdachtsdiagnose schließen läßt. Um hier trotzdem die Verdachtsgenerierung unabhängig von der Bildinterpretation bewerten zu können, werden im D3TRAINER folgende Schritte *vor* der eigentlichen Diagnosenkritik durchgeführt:

1. Überprüfe, ob falsche Bildinterpretationen die Systemdiagnosen beeinflußt haben. Dazu werden diese berichtigt und beobachtet, ob sich daraufhin die Systemdiagnosen ändern.

2. Unabhängig von einer möglichen Änderung werden als Basis der Bewertung die Menge von Systemdiagnosen benutzt, die aus den korrekten Daten gefolgert wurden.

3. Wenn die fehlerhafte Bildinterpretation die Systemdiagnosen verändert hatte, dann wird in dem Feedbackfenster eine entsprechende Textpassage[2] angezeigt und folgende zusätzliche Optionen werden angeboten:

[2] *Ihre fehlerhafte Bildinterpretation hat Einfluß auf die Diagnosenstellung. Bitte berichtigen Sie ihre Daten in der Fallpräsentation.*

- *Bildkritik*: Der Benutzer kann sich hier anschauen, welche Symptome er falsch erkannt hat.
- *Bildinterpretation berichtigen*: Automatisches Korrigieren der Bildinterpretation.
- *Was wäre wenn?*: Nun wird die Diagnosenkritik neu berechnet, wobei diesmal die Diagnosen als Bewertungsgrundlage benutzt werden, die auf der fehlerhaften Bildinterpretation beruhen.

Neben der Nutzung der situationsbezogenen Systemdiagnosen, kann es didaktisch sinnvoll sein, die endgültigen Diagnosen des Falles als Vergleichsgrundlage zu verwenden. Auch diese Option ist in dem D3TRAINER vorgesehen.

Nachdem jetzt die zu beachtenden Faktoren bei der Bewertung der Benutzerdiagnosen aufgezählt wurden, kann der vollständige Bewertungsalgorithmus hier angegeben werden. Der D3TRAINER bietet grundsätzlich zwei verschiedene Algorithmen an, bei denen **Algorithmus 1** eine genauere, aber auch strengere Bewertung errechnet, während **Algorithmus 2** einen mehr spielerischen Umgang bei der Diagnosenauswahl unterstützt.

Voraussetzungen:

- Menge an Systemdiagnosen S mit einem Systemstatus $s_s(s)$ (*verdächtigt* oder *bestätigt*) gemäß den Einstellungen situationsbezogen oder die Enddiagnosen des Falles.
- Menge an Benutzerdiagnosen D mit einem Sicherheitsfaktor $s_b(b)$ (*verdächtigt* oder *bestätigt*).

Von allen Systemdiagnosen $s \in S$ wird die Studentendiagnose $b \in D$ mittels der Funktion $B(s, D)$ herausgesucht, die sie am besten abdeckt. Anschließend wird der Ähnlichkeitsfaktor (s, b) errechnet und der fallbezogene Gewichtungsfaktor $G(s)$ einbezogen, um die Prozentzahl der erreichten Punkte gegenüber den erreichbaren Punkten für Systemdiagnosen und Benutzerdiagnosen zu errechnen.

Hilfsformeln:

(1)

$$B(s,D) = \begin{cases} v & : \exists v \in D \text{ Vorgänger und nicht} \\ & : \text{Nachfolger von } s \\ n & : \exists n \in D \text{ Nachfolger von } s \\ nil & : \text{sonst} \end{cases}$$

(2) Ähnlichkeit (s,b) nach Tabelle 3.1, $(s,nil)=0$

(3) Wertungsfaktor

$$W(d) = \begin{cases} 1 & : s_s(d) = \text{bestätigt} \\ w & : s_s(d) = \text{verdächtigt} \\ 1 & : \text{sonst} \end{cases}$$

mit w durch Wissensbasisentwickler vorgegebener Faktor (Default $w=0.5$)

(4) Fallbezogene Gewichtung $G(s)$ wie durch Experten eingegeben, Default $G(d)=1$

Formel für Algorithmus 1:

$$\text{Punkte} \quad P = \frac{\sum_{s\in S}(s,B(s,D))W(s)G(s)}{\sum_{d\in S\cup D}W(d)G(d)+\sum_{d\in S\vdash D}G(d)}$$

Für Algorithmus 2 wird getrennt betrachtet, wie gut die Systemdiagnosen durch die Benutzerdiagnosen abgedeckt werden (P_S) und wie gut die Benutzerdiagnosen an sich sind (P_B). Damit abgefangen wird, daß der Student viele Diagnosen angibt und somit einen guten Wert für P_S anstrebt und dabei keine Punkt für P_B riskiert, wird dieser Wert (P_B) nochmals auf ein Intervall zwischen $-0,5$ und 1 abgebildet. Aus den beiden Werten P_S und P_B wird ein Zahlenwert zwischen $-0,5$ und 1 gebildet, wobei P_S dreimal so stark gewichtet wird wie P_B.

Formeln für Algorithmus 2:

(1)

$$P_S = \frac{\sum_{s \in S}(s, B(s,D))W(s)G(s)}{\sum_{s \in S} G(s)}$$

(2)

$$P_B = \frac{\sum_{b \in D}(b)G(b)}{\sum_{b \in D} G(b)}$$

(3)

$$P = \frac{1}{4}(3P_S + (1,5P_B - 0,5))$$

Der entstehende Wert P wird dann benutzt, um die Leistung des Studenten bei der Auswahl von Verdachtsdiagnosen auf verschiedene Kategorien abzubilden. Diese Kategorien werden durch drei einheitliche Bilder (Applaus, Mittel, Daumen nach unten) in allen Feedbackfenster dargestellt. Bei der Bewertung der Diagnosenauswahl werden die Kategorien wie folgt bestimmt:

P **aus Algorithmus 1 oder 2**	**Kategorie**	**Bild**
$P \geq 0.75$	gut	Applaus
$0.25 < P < 0.75$	mittel	Mittel
$P \leq 0.25$	schlecht	Daumen nach unten

Diese Kategorien wurden gewählt, damit die Lernenden sich nicht durch Zahlenwerte ablenken lassen, und nur mit den groben Bewertungen umgehen, die sie auch schnell in dem Feedbackfenster erkennen können. Somit können sie sich problemlos einordnen und es entsteht ein spielerischer Wettkampf, das Applaus–Bildchen zu bekommen.

Eine Option, die im D3TRAINER noch nicht verfügbar ist, die aber durch das benutzte Wissensmodell grundsätzlich möglich wäre, ist, Konsequenzen einer Diagnose bei der Bewertung der Verdachtsdiagnosen des Studenten zu berücksichtigen. Das Wissensmodell muß nur um ein Attribut (z. B. *Risiko der Therapie*, *Gefährlichkeit der Diagnose* oder *Dringlichkeit der Therapie*) erweitert werden, das dann bei der Bewertung aber einfließt, daß bei nicht korrekten Diagnosen und Therapien ungefährlichere Fehler weniger stark gewertet werden. Aber auch mit dem jetzt vorhandenen Möglichkeiten können nahezu alle didaktischen Varianten bezüglich der Diagnoseauswahl und –bewertung in einer Fallbesprechung durch den D3TRAINER umgesetzt werden.

3.6.2 Selektion von Therapien

Die Entscheidung, ob und wie Therapien in dem System angeboten werden, ist dabei eine der grundlegendsten Varianten. Um Therapiebestimmung als diagnostische Option anbieten zu können, müssen diese natürlich in der D3 Wissensbasis modelliert worden sein. Aufgrund der hohen Ähnlichkeit zu den Diagnosen wurden Therapien als spezielle Diagnosen eingeführt, die auch in der Diagnosenhierarchie deklariert werden und deren Herleitung über die selben Mechanismen und Editoren eingetragen wird. Sind also Therapien in der Wissensbasis vorhanden, kann der Fachbereichsexperte zwischen mehreren Optionen wählen:

1. Therapien werden nicht angezeigt, d. h. für eine einfache Variante der Fallösung werden sie vollkommen ausgefiltert.
2. Therapien werden wie Diagnosen behandelt, d. h. sie werden mit den Diagnosen sowohl gemeinsam ausgewählt als auch gemeinsam bewertet.
3. Therapien werden bei jeder Diagnosenauswahl in der gleichen Hierarchie angezeigt und können hier ausgewählt werden, es findet aber eine getrennte Diagnosen– und Therapienbewertung statt.
4. Nach jeder Diagnosenauswahl und –bewertung müssen anschließend Therapien in einer eigenen Auswahlhierarchie selektiert werden, die dann auch separat bewertet wird.
5. Die Therapieauswahl und –bewertung findet vollkommen getrennt und auf Anfrage des Studenten statt.
6. Die Therapien werden wie zuvor nur optional angeboten, müssen jedoch bei der finalen Bewertung durch den Studenten angegeben werden.

Bei der Therapieauswahl und –bewertung stehen im wesentlichen die gleichen Varianten in der didaktischen Strategie zur Verfügung wie bei den Diagnosen, so daß die bei der Bewertung der Diagnosenauswahl genannten Algorithmen auch bei den Therapien verwendet werden. Auch bei der Begründung von Therapien stehen analoge Verarbeitungsweisen zur Verfügung, so daß im Folgenden nur noch von Diagnosen gesprochen wird, wobei die Therapien immer inbegriffen sind.

3.6.3 Begründung einer Verdachtsdiagnose

Das Begründen einer Verdachtsdiagnose in dem Expertensystem wird durch die Angabe der gefeuerten Regeln und deren Wahrscheinlichkeiten

umgesetzt. Soll also nun der Lernende seine Verdachtsdiagnose selbst begründen, müßte er in einem naiven Ansatz selbst Regeln aufstellen, die dann mit denen des Systems verglichen werden. Da Regeln aber oft auf unterschiedliche Weise formuliert werden und auch nicht das Erlernen der Expertenregeln sondern die Problemlösefähigkeit im Vordergrund stehen, kann dieser Lösungsansatz nicht benutzt werden. Eine Möglichkeit ist es, alle gefeuerten Regeln auf eine Symptome–Diagnosen–Relation zu abstrahieren, wobei noch beachtet werden muß, ob der diagnostische Mittelbau einbezogen oder ebenfalls aufgelöst werden soll.

Ähnlich wie bei der Formulierung der Regeln im Allgemeinen gibt es auch bei der Modellierung des diagnostischen Mittelbaus viel Spielraum für den Entwickler. Auch hier möchte man aus inhaltlichen Gründen auf eine Normung verzichten, so daß der Lernende nur die Symptome zur Begründung benützen kann, die er auch in der Fallpräsentation dargestellt bekommt.

Nach der Bewertung seiner Verdachtsdiagnosen kann oder, beim Modus *Symptomanforderung in Gruppen* nach der letzten Aufwandsgruppe, muß der Student eine von ihm gewählte Diagnose durch Angabe von Symptomen begründen. Die von ihm genannten Symptomausprägungen werden mit den echten hinweisenden Symptomen verglichen und es wird eine Bewertung daraus berechnet. Die einfachste Lösung wäre auch hier ein reiner Abgleich der beiden Symptommengen. Allerdings bringt alleine schon das Extrahieren der Symptome aus dem mehrstufigen heuristischen Herleitungsbaum einige Probleme mit sich. Geht man davon aus, daß eine Diagnose über mehrere Regeln hergeleitet wird, die unterschiedlich gewichtet werden und die wiederum über verschiedene Symptomabstraktionen hergeleitet werden, bleibt die Frage, wie man die Gewichtung der Regel auf die einzelnen Symptome verteilt. Sprechen nicht auch Symptome von Regeln für die Diagnose, wenn diese Regel aktuell nicht gefeuert hat? Versucht man den diagnostischen Mittelbau vor dem Studenten zu verstecken oder versucht man ihm diesen zu vermitteln? Das alles sind Fragen, die bei der Bereitstellung einer Begründung für die Verdachtsdiagnosen berücksichtigt werden sollten. Der D3TRAINER hat nicht alle Probleme lösen können, bietet aber eine variable Umgebung an, in der die meisten Probleme umgangen werden können.

Grundsätzlich kann der Wissensbasisentwickler entscheiden, ob er zur Begründung der Verdachtsdiagnosen Bezug auf die heuristischen oder die überdeckenden Regeln nehmen möchte. In beiden Fällen werden die Regeln dem Studenten nicht direkt vermittelt, sondern eine Symptom–Diagnosen–Relation extrahiert. Bei den heuristischen Regeln ist dies eine etwas schwierigere Aufgabe, da die Regelvorbedingungen oft sehr komplex und vielschichtig sind. Im wesentlichen wird hier die transitive Hülle über die erfüllten Vorbedingungen erstellt, während die Regelbewertungen auf diese verteilt werden. So erhält man zu einer Diagnose eine Menge an Ein-

zelkonditionen (beispielsweise *Allgemeinzustand* = *leichtes Fieber*) mit einem Bewertungsanteil. Durch diesen Bewertungsanteil können die Symptome (Einzelkonditionen, Symptomausprägungen) in verschiedene Kategorien eingeteilt werden:

- Stark dafür sprechende Symptome,
- dafür sprechende Symptome,
- schwach dafür sprechende Symptome und
- dagegen sprechende Symptome

Die Qualität dieser Aufteilung ist stark abhängig von der Wissensbasis, so daß der Fachexperte auch entscheiden kann, auf diese Einteilung zu verzichten und nur in dafür und dagegen sprechende Symptome aufzuteilen.

Bei der Benutzung der überdeckenden Regeln wird zwar derselbe Algorithmus verwendet, aber durch die grundsätzlich unterschiedliche Struktur (z. B. keine komplexen Vorbedingungen) können die wichtigen Symptome der Diagnose oft besser dargestellt werden. Auch läßt sich hier die Gewichtung meist leichter auf die einzelnen Symptome verteilen. Bei beiden Ansätzen ist man nun aber in der Lage, eine Menge von gewichteten Symptomen des Systems mit einer Menge von Symptomen des Studenten zu vergleichen.

Berücksichtigt man, daß in der Herleitung einer Diagnose auch eine andere Diagnose benutzt werden könnte, muß diese dem Lernenden natürlich auch zur Begründung zur Verfügung gestellt werden. Die Auswahl der Begründung zu einer Verdachtsdiagnose besteht also aus der Selektion von Symptomen und Diagnosen aus einer aufklappbaren Hierarchie, wobei ihm nur die Diagnosen zur Auswahl angeboten werden, die er zuvor als bestätigt deklariert hat.

Für die Bewertung werden nun die Symptome und Diagnosen des Studenten mit den aus den Regeln extrahierten Symptomen und Diagnosen gemäß des gewählten Problemlösers (heuristisch oder überdeckend) verglichen. Neben den bisher genannten Symptomklassen werden auch noch Symptome berücksichtigt, die in Regeln vorkommen, die grundsätzlich auf diese Diagnose zeigen, die aber aus unterschiedlichen Gründen nicht gefeuert haben. Die Menge dieser Symptome wird als *mit der Diagnose im Einklang stehend* bezeichnet. Auf Basis der errechneten Punktzahl wird die Leistung des Studenten in eine Kategorie eingeordnet und eine entsprechende Bewertung ausgegeben (siehe Abb. 3.28).

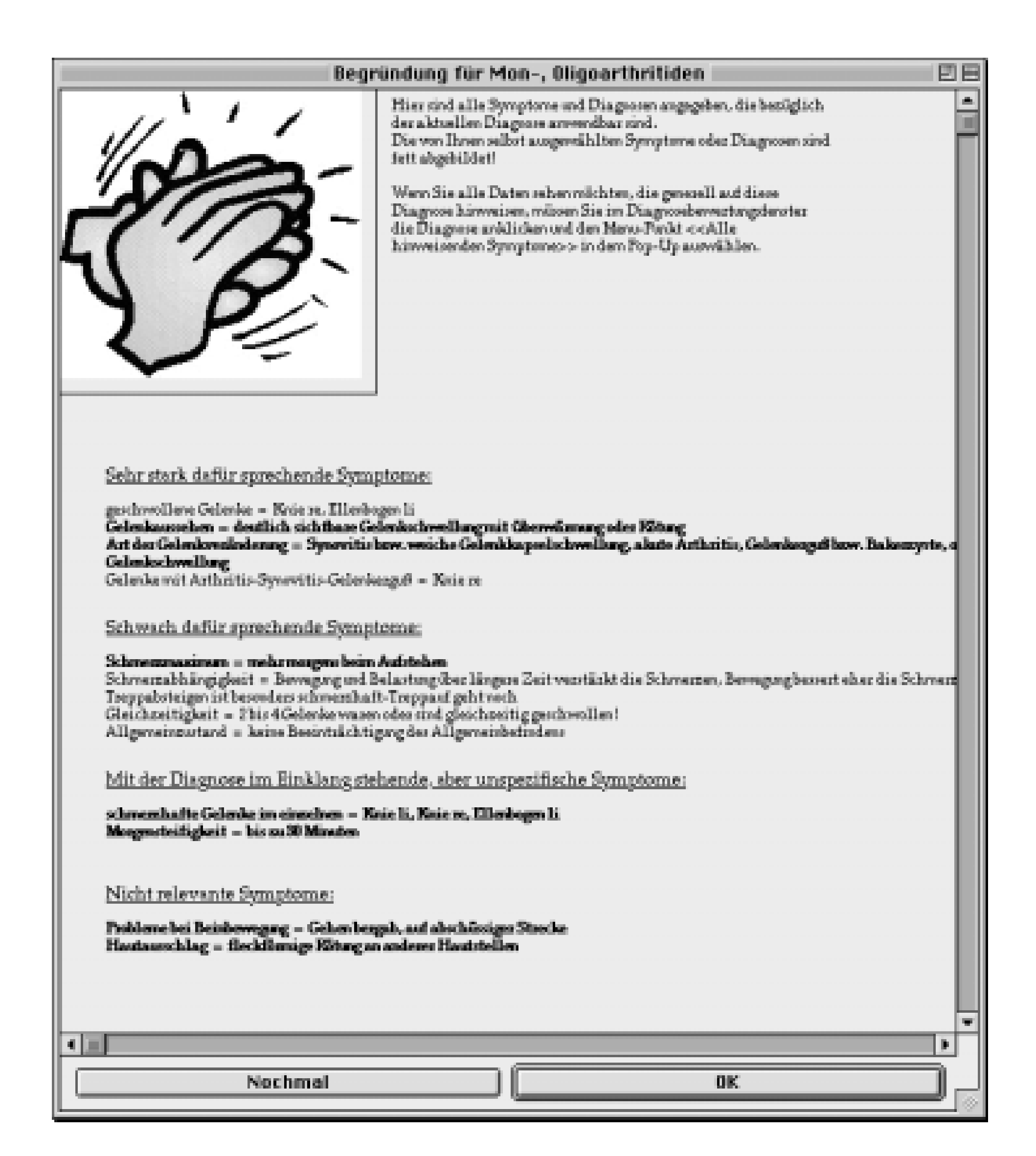


Abbildung 3.28: Feedbackfenster zur Bewertung der Begründung eines Studenten.

Der genaue Bewertungsalgorithmus arbeitet wie folgt:

Voraussetzungen zu einer Diagnose d:

- Eine Menge von extrahierten Diagnosen D und Symptomen S gemäß des gewählten Problemlösers (heuristisch, überdeckend), die in folgende Kategorien aufgeteilt werden:
  - S_{+++}: stark dafür sprechende Symptome,
  - S_{++}: dafür sprechende Symptome,
  - S_{+}: schwach dafür sprechende Symptome,
  - S_{-}: dagegen sprechende Symptome.
- Eine Menge von extrahierten Symptomen S_{ok}, die aus nicht gefeuerten Regeln extrahiert wurden. Sie stehen im Einklang mit der Diagnose.
- Eine Menge an Diagnosen und Symptomen B, die durch den Lernenden ausgewählt wurden.

Gewichtungsfunktion:
Gemäß der Zugehörigkeit zu einer Gruppe werden alle durch den Benutzer ausgewählten Objekte gewichtet:

$$G(s) = \begin{cases} 10 & : \quad s \in S_{+++} \\ 5 & : \quad s \in S_{++} \\ 3 & : \quad s \in S_{+} \\ 3 & : \quad s \in D \\ -5 & : \quad s \in S_{-} \\ 1 & : \quad s \in S_{ok} \\ 0 & : \quad \text{sonst} \end{cases}$$

Algorithmus zur Berechnung der Punkte:

$$\text{Benutzerpunkte} \quad P = \sum_{s \in B} G(s)$$

$$\text{Maximalpunkte} \quad M = \sum_{s \in S \cup D} G(s)$$

$$\text{Punktedifferenz} \quad \delta = M - P$$

$$Punkte = \begin{cases} 1 & : \ (P > 15) \wedge (\delta < 7) \\ 0.5 & : \ (P > 7) \wedge \delta beliebig \\ 0 & : \ \text{sonst} \end{cases}$$

Auch hier wird aufgrund der Gesamtbewertung die Aktion in eine Kategorie eingeordnet:

Gesamtbewertung P	**Kategorie**	**Bild**
1	gut	Applaus
0.5	mittel	Mittel
0	schlecht	Daumen nach unten

Ein wichtiger Aspekt bei dieser Bewertung ist die Tatsache, daß die Regeln, auf Basis derer die hinweisenden Symptome berechnet werden, nicht verschiedene Typen von Vorbedingungen vermischen. In der Expertensystemshell D3 können bei einer Regel Vorbedingung, Ausnahmen und Kontexte angegeben werden. So ist sichergestellt, daß bei den hinweisenden Symptomen wirklich nur diagnostische Hinweise und keine Randbedingungen erscheinen. Dies war eines der Probleme bei dem Tutorsystem Guidon, welches auf dem Expertensystem Mycin beruht. Da in Mycin Randbedingungen, wie beispielsweise das Alter des Patienten bei einer alkoholbedingten Infektionskrankheit, mit den wirklichen diagnostischen Hinweisen auf eine Stufe gestellt wurden.

Aufgrund des dynamischen Problemlösewissens und des statischen Erklärungswissens der Wissensbasis besitzt der D3TRAINER eine große **Erklärungsfähigkeit** in verschiedenen Bereichen der Fallbearbeitung. Hier kann zwischen fallbezogenen Hilfen und allgemeinen Informationen zu Diagnosen und Symptomen unterschieden werden. Fallbezogene Hilfen (Entscheidungshilfen) werden dem Studenten bei jeder Kritik seiner Aktion angeboten, um ihm zu helfen, die richtige Lösung zu erkennen oder zu sehen, warum seine Lösung nicht optimal ist. Im D3TRAINER werden folgende Entscheidungshilfen angeboten, wobei die mit $\star$ gekennzeichneten Punkte zu der *erweiterten Erklärung* gehören und auf Wunsch des Wissensbasisentwicklers ausgeblendet werden können:

Systemdiagnosen: Bei jeder Bewertung von Verdachtsdiagnosen des Studenten kann sich dieser die Diagnosen des Systems ansehen, mit denen seine Diagnosen verglichen wurden.

Hinweisende Symptome des Falles: Sowohl bei den korrekten Systemdiagnosen als auch bei seinen eigenen Verdachtsdiagnosen kann der

Student anfragen, welche Symptome des aktuellen Falles auf diese Diagnosen zeigen. Diese werden dann noch in verschiedene Kategorien (z. B. *stark dafür*, *im Einklang stehend*, *dagegen*) eingeteilt.

Alle hinweisenden Symptome: Zu jeder Diagnose kann sich der Student ansehen, welche Symptome generell, also unabhängig von dem aktuellen Fall, auf diese Diagnose zeigen. Dazu wird bei jeder Symptomausprägung angegeben, welchen Wert sie in dem aktuellen Fall hat.

⋆**Konkrete Herleitung:** Bei dieser Erklärungsoption werden die der Herleitung zugrundeliegenden Regeln und das genaue Punkteschema der internen Bewertung angezeigt. Diese Option, wie auch die beiden folgenden, sind sehr stark an die Wissensrepräsentation des Expertensystems angebunden und zeigen auch die nicht abstrahierten Regeln in einer *Wenn–Dann*–Darstellung. Diese Details sind nicht für alle Trainingssysteme sinnvoll und können deshalb über die Präferenzen des D3TRAINERS ausgeblendet werden.

⋆**Herleitungsgraph:** Darstellung der gefeuerten Regeln in einem Graph mit dem gesamten diagnostischen Mittelbau.

⋆**Warum nicht?** Bei Diagnosen, die durch das System nicht ausgewählt wurden, kann man sich über diese Option ansehen, warum die entsprechenden Regeln des Systems nicht gefeuert haben. Mit dieser Option kann der Student dann sehen, welche Voraussetzungen fehlen, ob Ausnahmen zutreffen oder ein gewisser Kontext für diese Diagnose nicht gegeben ist.

Durch die Verbindung von hoher Erklärungsfähigkeit und detaillierten Kritikmöglichkeiten bei der Bewertung von Verdachtsdiagnosen und ihren Begründungen bieter der D3TRAINER eine besonders gute Lernumgebung für diesen Bereich der Fallbesprechung.

3.6.4 Selektion von neu anzufordernden Informationen

Der D3TRAINER bietet verschiedene didaktische Varianten der Symptomanforderung an, bei denen der Student selbst die Auswahl neuer Informationen steuert. In diesem Fall kann die Selektion auch durch das System bewertet werden. Wie auch bei den Diagnosen können die Frageklassen entweder in einer multiple–choice Auswahl selektiert werden oder über die Gesamthierarchie aller in der Wissensbasis möglichen Optionen. Zur Erklärung (Entscheidungshilfe) bei der Anforderung neuer Informationen kann der Student abfragen, welche Untersuchung das System als nächstes

vorschlagen würde. Er kann sich dabei entscheiden, ob er diese Untersuchung durchführen oder seine eigene Wahl treffen möchte.

Grundlage bei der Bewertung von Untersuchungsanforderungen sind zwei Größen: der Informationsgewinn, den man bei der Fallbearbeitung durch die Kenntnis der neuen Information erlangt, und der Aufwand, der entstünde, wenn diese Untersuchungen in der Realität durchgeführt werden würden. Eine sinnvolle Selektion neuer Informationen würde dann bedeuten, daß immer die Untersuchung angefordert werden würde, deren Kosten–Nutzen–Abschätzung am besten erscheint. In manchen Domänen gibt es zusätzlich noch gewisse Abarbeitungsreihenfolgen, die sich standardmäßig entwickelt haben, so z. B. in der Medizin die Reihenfolge *Anamnese, körperliche Untersuchung, Labor und technische Untersuchungen*.

Die Bewertung der angeforderten Untersuchungen des Studenten erfolgt also durch Vergleich mit der Untersuchung, die das System als nächstes vorschlagen würde. Die Indizierung kann dabei entweder direkt durch Regeln, zur Klärung von Verdachtsdiagnosen oder über eine Kosten–Nutzen–Relation erfolgt sein. Da die Reihenfolge der Abarbeitung oftmals willkürlich ist, soll diese natürlich dem Lernenden nicht aufgezwungen werden, d. h. es können grundsätzlich alle Untersuchungen jederzeit gewählt und durchgeführt werden. Bei der Bewertung der ausgewählten Tests muß aber diese Freiheit in der Reihenfolge auch berücksichtigt werden. Manchmal kann es auch vorkommen, daß eine Untersuchung für den Fall unbedingt notwendig ist, in der momentanen Fallsituation aber durch das System nicht vorgesehen wird. Für diese Fälle werden in dem Formular zum Abspeichern eines Falles alle notwendigen Frageklassen zu dem Gesamtfall abgelegt, so daß bei der Bewertung darauf zurückgegriffen werden kann.

Der Algorithmus vergibt für jede der gewählten Frageklassen eine Punktzahl und bildet dann das Mittel über die gesamte Auswahl. Der Mittelwert wird dann für die Einordnung in eine der drei Kategorien verwendet (siehe Abb. 3.12).

Der genaue Bewertungsalgorithmus arbeitet wie folgt:

Voraussetzungen:

- Eine geordnete Liste I von Frageklassen, die durch das System direkt indiziert wurden.
- Eine geordnete Liste K von Frageklassen, die über die Kosten–Nutzen–Analyse indiziert wurden.
- Eine geordnete Liste N von Frageklassen, die beim Abspeichern des Falles als notwendige Frageklassen deklariert wurden.
- Eine Menge von Frageklassen F_B, die durch den Lernenden ausgewählt wurden.

Gewichtungsfunktion:
Gemäß der Zugehörigkeit zu einer Gruppe werden alle durch den Benutzer ausgewählten Frageklassen gewichtet:

$$G(f) = \begin{cases} 100 & : \; f \text{ erstes Element von } I \\ 50 & : \; f \in I \\ 50 & : \; f \in K \\ 50 & : \; f \in N \\ 0 & : \; \text{sonst} \end{cases}$$

Algorithmus zur Berechnung der Punkte:

$$\text{Mittelwert} \quad M = \frac{\sum_{f \in F_B} G(f)}{\sum_{f \in F_B} 1}$$

$$P = \begin{cases} 1 & : \; M \geq 50 \\ 0.5 & : \; 30 \leq M < 50 \\ 0 & : \; M < 30 \end{cases}$$

Auch hier wird aufgrund der Gesamtbewertung die Aktion in eine Kategorie eingeordnet:

Gesamtbewertung P	**Kategorie**	**Bild**
1	gut	Applaus
0.5	mittel	Mittel
0	schlecht	Daumen nach unten

Die Zuordnung von Punkten zu Kategorien stellt hier, wie in einigen anderen Bewertungsalgorithmen im D3TRAINER, eine Schätzung dar, die sich in einigen Tests als günstig erwiesen hat.

Die entstehenden Kosten dieser Untersuchungsauswahl werden in bis zu drei Posten (*Kosten*, *Risiko* und *Zeitaufwand*) aufgeteilt, die der Wissensbasisentwickler selbst festlegen und benennen kann. Diese Kosten werden in einer Balkendarstellung aufsummiert, wobei bei Bedarf eine Detailansicht aufgerufen werden kann, in der die Beiträge der einzelnen Untersuchungen zu der Gesamtsumme dargestellt werden (siehe Abb. 3.13). Diese Kosten spielen vor allem bei der Abschlußbewertung (siehe Abschnitt 3.6.7) eine Rolle.

3.6.5 Begründung der angeforderten Information

Das Diagnosesystem D3 verfügt wie bereits vorgestellt über mehrere Mechanismen zur Indikation von Untersuchungen. Der naive Ansatz zum

Vergleich von System– und Benutzeraktion wäre die Angabe dieser Mechanismen bei der Begründung einer Frageklasse. Die komplizierte Verrechnung ist für den Studenten aber oft nicht nachzuvollziehen, außerdem ist, wie schon erwähnt, die Abarbeitungsreihenfolge auch manchmal willkürlich. Deshalb bietet der D3TRAINER nur die Indikation über Verdachtsdiagnosen als Begründungsmöglichkeit für den Studenten an. Der Student kann seine Untersuchungen durch Angabe von Verdachtsdiagnosen begründen, die durch die neue Information entweder bestätigt oder ausgeschlossen werden sollen. Die Kritikkomponente überprüft nun erstens, ob die angegebenen Diagnosen durch die selektierten Frageklassen geklärt werden könnten und zweitens, ob diese Diagnosen überhaupt verdächtigt sind.

Der genaue Bewertungsalgorithmus arbeitet wie folgt:

Voraussetzungen zur Begründung der Frageklasse f:

- Eine Menge an Diagnosen $K(f)$, die grundsätzlich von der Frageklasse f geklärt werden könnten.
- Eine Menge an Diagnosen S, die gerade durch das System verdächtigt werden.
- Eine Menge an Verdachtsdiagnosen B, die durch den Lernenden ausgewählt wurden, und die durch die Frageklasse f geklärt werden sollen.

Algorithmus zur Bewertung:

$$P = \begin{cases} 1 & : \quad (K(f) \cap B) \neq \emptyset \wedge (B \cap S) \neq \emptyset \\ 0.5 & : \quad (K(f) \cap B) \neq \emptyset \\ 0 & : \quad sonst \end{cases}$$

Auch hier wird aufgrund der Gesamtbewertung die Aktion in eine Kategorie eingeordnet:

Gesamtbewertung P	**Kategorie**	**Bild**
1	gut	Applaus
0.5	mittel	Mittel
0	schlecht	Daumen nach unten

Der Grund für die Abstraktion von der internen Abarbeitung zu dieser Begründung liegt in der Tatsache, daß es sich bei den Indikationsregeln in der Wissensbasis um kompiliertes Wissen handelt. Auch diese Regeln sind

ja eigentlich dazu da, verdächtigte Diagnosen zu klären, gehen aber nicht den Schritt über Fakt–Diagnose–Frageklasse, sondern beschreiben gleich die Fakt–Frageklasse Relation. Dieses kompilierte Wissen ist deshalb didaktisch nicht sinnvoll und sollte zur Ausbildung nicht benutzt werden. Da man es aber aus der Konsultationswissensbasis nicht automatisch wieder extrahieren kann, muß es für einen didaktisch sinnvollen Einsatz zusätzlich durch den Wissensbasisentwickler modelliert werden. In den bisherigen Anwendungen ist dies durch die zusätzliche Eingabe von Regeln geschehen.

3.6.6 Erkennen von Merkmalen aus Bildern

Das Erkennen von Symptomen aus medialem Material ist eine der attraktivsten Erweiterungen, die aber nur durch zusätzliches Wissen zu der D3–Wissensbasis generiert werden kann. Durch Angabe eines Bildsystems zu einem Fall können Teile der Symptommenge aus einem Bild erkannt werden und müssen nicht durch das System in verbaler Form beschrieben werden. Neben der erhöhten Realitätstreue, die eine mediale Darstellung mit sich bringt, ermöglicht ein Bildsystem auch das Einüben von Bildinterpretationen, wie sie auch im Diagnosealltag vorkommen.

Ein Bildsystem, wie in Abschnitt 3.3.3 beschrieben, besteht aus mehreren Bildern, die eine interne Verknüpfung zu den Symptomen des Falles haben, die in diesem Bild erkannt werden können. Alle Bilder zu einem Symptom werden in einem Karteikasten (siehe Abb. 3.15) zusammengestellt, zwischen dessen Karten der Student wechseln kann. Der Übergang zu einem anderen Bild bewirkt, daß die linke Seite des Fensters so umgebaut wird, daß im oberen Bereich die zu erkennenden Symptome angezeigt werden, deren Antwortalternativen im unteren Bereich zur Verfügung gestellt werden. Wechselt der Student oben zwischen den Symptomen, so werden die entsprechenden Antwortalternativen im unteren Bereich aktualisiert. Durch Ausfüllen dieses Bereichs können nun die erkannten Symptombelegungen eingetragen werden. Die Kritik einer Bildinterpretation kann dabei entweder generell für den ganzen Fall oder aus einem bestimmten Bild heraus aufgerufen werden.

Der einfachste Ansatz ist wieder, die eingegebenen Symptomausprägungen mit denen zu vergleichen, die bei der Beschreibung der Bilder angegeben wurden. In vielen Bildern kommt es aber vor, daß manche Ausprägungen deutlicher zu sehen sind oder gewisse Ausprägungen grundsätzlich wichtiger sind. Hier ist es von Vorteil, wenn eine Gewichtung der Symptomausprägungen angegeben werden kann, mit der deren Wichtigkeit in Beziehung gesetzt werden kann. Diese Gewichtung kann im D3TRAINER entweder nur pro Bild oder für das ganze Bildsystem gelten. Durch die Gewichtung für den ganzen Fall können somit nicht nur Symptome innerhalb

eines Bildes sondern in dem gesamten Bildsystems untereinander in Bezug gesetzt werden.

Der Bewertungsalgorithmus unterteilt alle erkennbaren Symptome und die Symptomeingabe des Benutzers in richtige, falsche oder nicht erkannte Symptome. Nun werden die Gewichtungen noch berücksichtigt und der Quotient von richtig zu falsch oder nicht erkannten Symptomen berechnet, der dann für die Einteilung in die Kategorien benutzt wird (siehe Abb. 3.15).

Der genaue Bewertungsalgorithmus arbeitet wie folgt:

Voraussetzungen:

- Eine Menge B an Symptomen, die durch den Studenten eingegeben wurde.
- Eine Menge S an Symptomen, die für diesen Fall korrekt sind.
- Eine Menge $S_+ = S \cap B$ der richtig erkannten Symptome.
- Eine Menge $S_- = B \setminus S$ der falsch erkannten Symptome.
- Eine Menge $S_0 = S \setminus B$ der nicht erkannten Symptome.
- Eine Funktion $G(s)$, die die Gewichtung innerhalb der Bildsysteme bestimmt (je nach Einstellung fall– oder bildbezogen).

Algorithmus zur Berechnung der Punkte:

$$P = \frac{\sum_{s \in S_+} G(s)}{\sum_{s \in S} G(s)}$$

Auch hier wird aufgrund der Gesamtbewertung die Aktion in eine Kategorie eingeordnet:

Gesamtbewertung P	**Kategorie**	**Bild**
$P = 1$	optimal	Applaus
$1 > P \geq 0.75$	überwiegend richtig	Applaus
$0.75 > P \geq 0.50$	meistens richtig	Mittel
$0.5 > P \geq 0.25$	vieles richtig	Mittel
$0.25 > P \geq 0.001$	manches richtig	Daumen nach unten
0	alles falsch	Daumen nach unten

Als Entscheidungshilfe bei der Bildinterpretation kann der Benutzer sich zu jedem Bild die korrekte Lösung anzeigen lassen. Außerdem kann über

die Option *Wo hätte ich das sehen sollen?* aus dem Feedbackfenster die Stelle in dem Bild angezeigt werden, in der genau diese Symptomausprägung zu erkennen ist.

Eine weitere Möglichkeit, in den Bewertungsprozeß eingreifen zu können, wäre es, Ähnlichkeiten zwischen zwei Symptomausprägungen zu berücksichtigen, d. h. ähnlich wie bei den Ober– und Folgediagnosen bei der Verdachtsüberprüfung einen zusätzlichen Faktor zu berechnen, der auch noch Punkte vergibt, wenn man ein ähnliches Merkmal erkannt hat. Diese Option ist aber im D3TRAINER noch nicht umgesetzt. Auch eine Begründung der Bildinterpretation ist nicht umgesetzt und aus den in Abschnitt **??** genannten Gründen auch nicht geplant. Eine mögliche Umsetzung der Begründung für das Erkennen von Merkmalen in einem Bild ist aber die Eingabe eines erläuternden Textes, der zu jedem Bild erklärt, was wo und warum erkannt werden kann und wo bei diesem Bild die Besonderheiten sind. Diese Texte müssen dann bei der Erstellung der Bildsysteme in die Beschreibung der einzelnen Bilder durch Symptome des Expertensystems integriert werden.

3.6.7 Abschlußbewertung

Am Ende jeder Fallbearbeitung wird eine gemeinsame Abschlußbewertung über alle Bereiche der diagnostischen Problemlösung erstellt (siehe Abb. 3.16), die dem Studenten helfen soll, seine Fallbearbeitung insgesamt besser einzuschätzen. In der Abschlußkritik werden folgende Bereiche einzeln bewertet:

- Diagnoseauswahl,
- Therapieauswahl,
- Informationsanforderung,
- Kosten–Aufwand–Risiko und
- Bildinterpretation

Um dem Fachexperten die Möglichkeit zu bieten, seinem Trainingssystem einen Schwerpunkt bezüglich der verschiedenen Teilaufgaben der Fallbearbeitung zu gegeben, können für Diagnosenfindung, Therapienstellung und Umgang mit *Kosten–Aufwand–Zeit* Faktoren angegeben werden, die diese Teilaufgaben gegeneinander gewichten. Für die Diagnosen und Therapiebewertung werden Prozentzahlen berechnet, die ausdrücken, wie gut diese finale Lösung war. Dabei bezieht sich der Vergleich nicht mehr auf die aktuelle Fallsituation, sondern auf die echte Lösung des Falles. Mit den

einzelnen Bewertungen und den Gewichtungen wird nun eine Prozentzahl ermittelt, wie gut die Fallösung des Studenten ist. Die Kostenabschätzung führt noch zu einer Abwertung dieser Prozentzahl, wenn sie relevant höher ist, als der Fallaufwand, den der Bereichsexperte als sinnvoll bezeichnet hat. Die Bewertung der Frageklassen, sowie die Kritik der Bildinterpretionen werden zwar in dem Feedbackfenster angezeigt, gehen aber nicht direkt in die Fallbewertung ein. Grund dafür ist, daß angenommen wird, daß der Lernende schon sehr schnell die Fallösung erkennt und den Fall aus Zeitgründen nicht vollständig durchspielen möchte. Da die Eingabe der Bildinterpretationen sowie das Anfordern aller notwendigen Frageklassen sehr zeitintensiv sein können, werden die entsprechenden Bewertungen zwar gezeigt, wirken sich aber eben nicht negativ auf die Gesamtbewertung aus.

Auch in diesem Bewertungsfenster können alle Kritikpunkte genau erklärt werden, so daß der Student das gegebene Feedback schnell umsetzen kann. Von hier aus kann er auch wieder zurück in seine aktuelle Fallsituation wechseln und den Fall weiter bearbeiten.

Der genaue Bewertungsalgorithmus arbeitet wie folgt:

Voraussetzungen:

- Die von dem Benutzer als finale Fallösung bezeichneten Diagnosen und Therapien.
- Die während der Fallbearbeitung ausgewählten Frageklassen des Benutzers mit ihren Kosten, Risiken und dem entsprechenden Zeitaufwand.
- Eine Bewertung der Bildinterpretation.
- Alle Systeminformationen zu den notwendigen Frageklassen, deren Kosten, Risiken und Zeitaufwand, die korrekten Diagnosen und Therapien, sowie die korrekten Bildinterpretationen.
- Die durch den Wissensbasisentwickler festgelegten Gewichtungen für die Diagnosenbewertung F_D, die Therapiebewertung F_T und die Kostenbewertung $F_\$$.

Hilfsformel für die Abwertung durch die Kosten:

$$\text{maximaler Überschuß} \quad M = \max \begin{cases} \frac{K}{K\star} \\ \frac{R}{R\star} \\ \frac{Z}{Z\star} \end{cases}$$

mit K reale Kosten und $K\star$ maximal vorgesehene Kosten, R reale Risiken und $R\star$ maximal vorgesehene Risiken und Z realer Zeitaufwand und $Z\star$ maximal vorgesehener Zeitaufwand.
Der Abwertungsfaktor A_K wird wie folgt berechnet:

$$A_K = \begin{cases} 0 & : \; M \leq 0 \\ 0.01M & : \; M < 1.5 \\ 0.01M - 0.05 & : \; M < 2 \\ 0.05M - 0.05 & : \; M < 5 \\ 0.02M - 0.2 & : \; M < 10 \\ 0.5 & : \; sonst \end{cases}$$

Die Bewertungen für Diagnosen B_D und Therapien B_T wurden durch die zuvor beschriebenen Formeln berechnet, die Faktoren F_D und F_T wurden durch den Benutzer oder den Experten festgelegt.

Algorithmus zur Berechnung der Gesamtpunktzahl:

$$P = \frac{F_D B_D + F_T B_T}{F_D + F_T} - A_K$$

Auch hier wird aufgrund der Gesamtbewertung die Aktion in eine Kategorie eingeordnet:

Gesamtbewertung P	**Kategorie**	**Bild**
$P = 1$	perfekt	Applaus
$1 > P \geq 0.90$	sehr gut	Applaus
$0.90 > P \geq 0.75$	gut	Mittel
$0.75 > P \geq 0.25$	durchschnittlich	Mittel
$P < 0.25$	nicht gelöst	Daumen nach unten

Nach der Abschlußkritik des Falles wird noch ein Expertenkommentar gezeigt, in dem der Fallautor Besonderheiten des Falles nochmals hervorheben und abschließende Bemerkungen machen kann (siehe Abb. 3.17).

Während der gesamten Fallbesprechung wurden alle Aktionen des Studenten protokolliert und die Bewertungen seiner Aktionen gespeichert. Dieses *Fallverlaufsprotokoll* kann am Ende des Falles gesichert werden, um eine

spätere Auswertung vorzunehmen. Es existiert auch die Möglichkeit, diese Protokolle automatisch abzuspeichern, was besonders bei einer Feldstudie zur Evaluation sehr nützlich ist.

3.7 Zusammenfassung

Der D3TRAINER bietet, wie in diesem Kapitel gezeigt, fast alle didaktische Variationen und Kritikoptionen an, die in Abschnitt **??** vorgestellt wurden. Ausnahmen sind dabei nur die natürlichsprachliche Verbalisierung von Symptomen und die Eingabe mit Kritik von Begründungen zu den Bildinterpretationen. Die Umwandlung der *Symptom–Wert*–Verbalisierungen in natürlichsprachliche Sätze könnte technisch gesehen ohne größere Probleme in den D3TRAINER integriert werden, worauf aber zugunsten der übersichtlichen Darstellung des Gesamtfalles verzichtet wird. Eine Eingabe zur Begründung der Bildinterpretationen und deren Kritik durch das System stellt für Student und Fachbereichsexperten einen so großen Aufwand dar, daß auch auf diese Funktionalität verzichtet wird, da die entstehenden Kosten den geringen tutoriellen Nutzen weit übersteigen.

Durch die Einstellungsmöglichkeiten in dem D3TRAINER kann nicht nur der Fachbereichsexperte dem entstehenden Trainingssystem eine persönliche Note geben und seinen bevorzugten Lehrstil anbieten, sondern auch der Student kann bezüglich seinen Lernvorausetzungen eine geeignete Variante wählen. So können fortgeschrittene Fachärzte mit geringem Zeitpotential mit dem selben System arbeiten wie Studenten, die sich gerade erst in dieses Gebiet einarbeiten und viel Unterstützung benötigen. Gerade bei diesen beiden Extremen spielt auch das Angebot an Fällen eine große Rolle, stellt man doch dem Anfänger typische einfache Fälle zur Verfügung, während der Facharzt an exotischen schwierigeren Fällen interessiert ist. Für die Bereitstellung verschiedener Benutzerprofile können verschiedene Einstellungen als didaktische Modi definiert und mit einem Namen versehen werden, so daß diese didaktisch sinnvollen Kombinationen über einen Menüeintrag zur Verfügung stehen.

Wenn auch in dieser Arbeit meist mit medizinischen Beispielen gearbeitet wird, darf nicht vergessen werden, daß es sich um ein universelles System für das Erlernen diagnostischer Problemlösung handelt. Zwar stammen die meisten fertiggestellten Anwendungen des D3TRAINERS aus der Medizin, es wurde aber während der gesamten Entwicklungsphase Wert auf die Offenheit bezüglich der allgemeinen Einsetzbarkeit gelegt.

4

Evaluation

Die Verwendung eines Computerprogramms in der Ausbildung impliziert auch eine Verantwortung des Entwicklungsteams für die Qualität dieser Lehre. Somit ist Qualitätssicherung ein wichtiger Bestandteil bei der Entwicklung solcher Systeme und muß in den gesamten Entstehungsprozeß intergriert werden. Gerade bei der Vielzahl der im Moment vorhandenen CBT–Systeme helfen gemeinsame Qualitätskriterien dem Lernenden, einen Überblick über die vorhandenen Möglichkeiten zu bekommen.

4.1 Grundlagen der Evaluation

Der Begriff *Evaluation* wird in vielen verschiedenen Fachrichtungen wie beispielsweise der Psychologie, der Informatik und der Pädagogik benutzt. Für die weitere Verwendung des Begriffes und der unterschiedlichen Formen und Methoden, die zur Evaluation von Lernsystemen benutzt werden, wird in diesem Kapitel eine Begriffsklärung und eine Aufzählung der möglichen Methodiken vorgenommen.

4.1.1 Evaluation – Begriff und Formen

Nach dem Wörterbuch der Psychologie [Psycho 1990] wird Evaluation wie folgt definiert:

> „Evaluation [von engl. value „Wert“]: die Auswertung einer Erfahrung durch eine oder mehrere Personen; der Begriff ist bes. gebräuchlich in den Sozialwiss., die soziale oder pädagog. Aktionsprogramme auf den in ihnen angestrebten Erfolg hin untersuchen.“

Überträgt man diese Begriffsbestimmung auf die Evaluation von Lernsoftware, so soll diese also den Erfolg eines Lernprozesses messen. Für die Evaluation intelligenter Trainingssysteme werden verschiedene Formen (siehe [MARK and GREER 1993], [BLOCK 1996]) vorgeschlagen, in die man die Evaluationen einteilen kann. Eine grundsätzliche Entscheidung betrifft den Zeitpunkt der Evaluation bezüglich des Entwicklungsstandes des Systems. Findet sie entwicklungsbegleitend statt, spricht man von einer *formativen* Evaluation. Diese wird benutzt, um schon zu einem frühen Zeitpunkt korrigierend in die Entwicklung eingreifen zu können und etwaige Probleme zu beseitigen. Findet die Evaluation am fertigen Produkt statt, wird eine *summative* Evaluation durchgeführt, die eine reine Qualitätsbewertung des Programms darstellt und dessen Erkenntnisse erst in einer späteren Weiterentwicklung genutzt werden.

Die Unterscheidung in *Laborevaluationen* und Evaluation im *realen Einsatz* hängt meist mit dem Entwicklungsstand des Systems zusammen. Für formative Evaluationen wird oft eine künstliche Umgebung (Labor) geschaffen, in der ausgewählte Teilnehmer mit dem System arbeiten. Der Vorteil dieser Art der Evaluation ist, daß viele Variablen des Lernprozesses betrachtet werden können und weitgehend unter Kontrolle der Untersuchenden sind. Für eine summative Evaluation ist jedoch ein Einsatz in der realen Umgebung unerlässlich, so soll schließlich der Wert des Programms in seiner endgültigen Anwendungssituation bestimmt werden.

Desweiteren kann zwischen einer *intrinsischen, Ergebnis– oder Kontextevaluation* unterschieden werden. Die intrinsische Evaluation beurteilt die Übereinstimmung von Zielen, Inhalt und Methoden der Software. Bei der Ergebnisevaluation werden ausschließlich die Auswirkungen, beispielsweise die Lerneffizienz und der Leistungszuwachs, bewertet, während bei der Kontextevaluation die pädagogischen Voraussetzungen betrachtet werden. Hier liegt der Untersuchungsmittelpunkt auf der gegenseitigen Beeinflussung verschiedener Unterrichtsfaktoren.

Die Tatsache, wer die Evaluation durchführt, wird bei der Einteilung in *äußere* und *innere* Evaluation betrachtet. Leitet einer der Entwickler oder Autoren des Systems selbst die Studie, so spricht man von einer inneren Evaluation, während das Entwicklungsteam an einer äußeren Evaluation nicht beteiligt ist. Bei einer *Mikroevaluation* stehen bestimmte Teile des Gesamtsystems, beispielsweise nur die Lehrbuchoption, im Mittelpunkt, während bei der *Makroevaluation* das Gesamtsystem als Einheit betrachtet wird.

Kann man in der Evaluation das System mit einem anderen System oder einer anderen konventionellen Lehrmethode vergleichen, spricht man im Gegensatz zu der *nicht–vergleichenden* von einer *vergleichenden* Evaluation. Unterteilt man die Gesamtheit aller Lernenden in zwei Gruppen und unterrichtet nur eine Gruppe mit dem System, nennt man die alternativ unterrichtete Gruppe *Kontrollgruppe*. Oft ist das Finden einer Kontrollgrup-

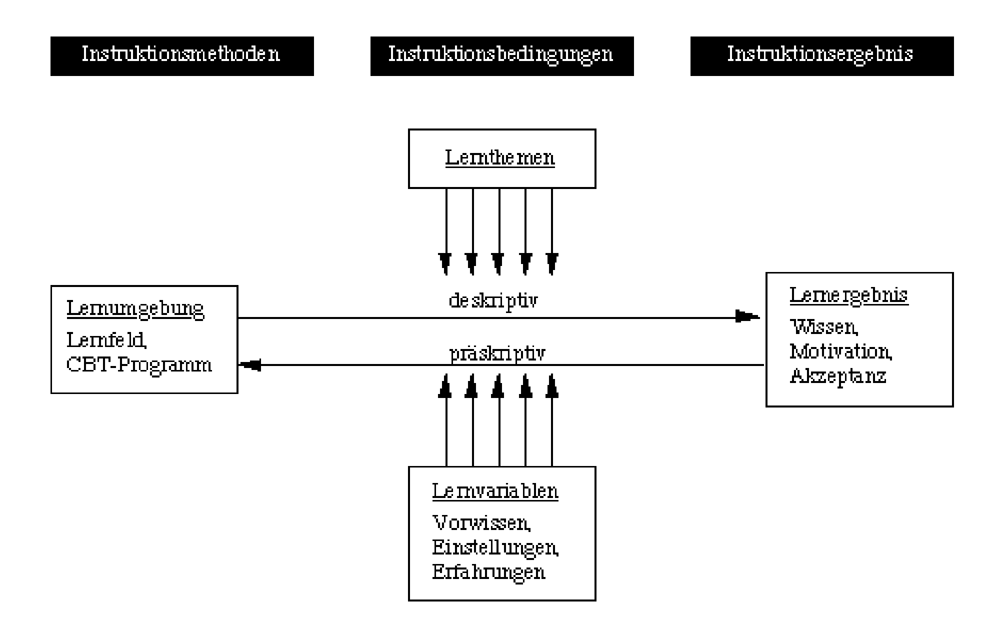


Abbildung 4.1: Evaluationsmodell nach einem theoretischen Ansatz des Instruktionsdesign aus [AUHUBER 1998].

pe schwierig, da sich die Gruppen in den sonstigen Lernvoraussetzungen nicht wesentlich unterscheiden sollen, damit dadurch das Untersuchungsergebnis nicht verfälscht wird. Findet die Evaluation nicht in einer Laborumgebung sondern in einem realen Einsatz statt, kommt noch hinzu, daß man allen Lernenden dieselben Lernmöglichkeiten bieten muß. Der Entzug des Lernprogramms für die eine Gruppe würde einer Diskriminierung entsprechen. Eine der Möglichkeiten, dieses Problem zu vermeiden, ist der *Crossover–Versuchsplan* [SCHULZ et al. 1997]. Hier bekommen zwei Lernergruppen 1 und 2 für zwei verschiedene Lerninhalte A und B abwechselnd einmal ein Computerprogramm und einmal ein herkömmliches Lehrbuch zur Verfügung gestellt.

In [AUHUBER 1998] wird ein Evaluationsmodell nach einem theoretischen Ansatz des Instruktionsdesigns vorgestellt (siehe Abb. 4.1), welches Instruktionsmethoden, Instruktionsbedingungen und Instruktionsergebnisse betrachtet. Die Instruktionsmethoden beschreiben die gesamte Lernumgebung, also das Lernfeld sowie das CBT–Programm. Die Lernthemen und die Lernvariablen Vorwissen, Einstellungen und Erfahrungen sind die Instruktionsbedingungen. Wissen und Motivation der Lernenden und ihre Akzeptanz des Systems gehören zu dem Lernergebnis, das am Ende der Studie betrachtet werden kann.

4.1.2 Methoden der Evaluation

Abhängig von der Entscheidung für eine Form der Evaluation kann nun die Studie genauer geplant werden. Hierfür stehen verschiedene Methoden zur Verfügung, die eingesetzt werden können, wobei nicht jede Methode für jede Untersuchung sinnvoll ist. Viele der vorgestellten Methoden kommen aus der Instruktionspsychologie, andere aus der Künstlichen Intelligenz zur Bewertung von Wissensmodellen bei Expertensystemen.

4.1.2.1 Studienunabhängige Methoden

Die Methoden dieser Art benötigen für ihre Durchführung keine Gruppe an Lernenden, die mit dem System kontrolliert arbeiten. Die Bewertung findet anhand anderer Kriterien und Schwerpunkte statt. Die erste Methode dieser Evaluationsklasse ist die *Expertenbefragung*, d. h. man bittet ‚menschliche' Experten des Fachgebiets, das System zu testen und bezüglich der gewünschten Aspekte zu bewerten. Diese Methode eignet sich besonders gut für die formative Laborevaluation, da man die Hinweise und Ratschläge der versierten Experten noch sehr gut in die weitere Entwicklung einbauen kann. Ein Vorteil davon ist auch, daß ein späterer Einsatz durch diese Fachexperten erleichtert wird, da im allgemeinen ihre Identifikation mit dem System steigt, wenn sie an der Entwicklung beteiligt sind.

Systemen mit internem Wissensmodell wie intelligente Tutorsysteme steht eine formale Methode zur Verfügung, wie dieses Wissensmodell zu bewerten ist. Durch *Validierung* und *Verifizierung* des Wissensmodell kann somit auch eine Aussage über die Qualität des Gesamtsystems gemacht werden. Schließlich ist ein Intelligentes Tutorsystem immer nur so gut, wie das Wissen, das es vermittelt. Validierung heißt in diesem Umfeld, ob das System das tut, was es tun soll, während Verifizierung zu der Aussage führt, ob es dies richtig tut. Auf die genauen formalen Methoden zu Validierung und Verifizierung wird hier nicht genauer eingegangen, da sie stark von dem verwendeten Wissensmodell abhängen (siehe [BALCI 1997]). Bei der Verwendung von Wissensbasen eines Expertensystems zur Diagnostik kann dies beispielsweise der Testlauf mit vielen Fällen mit bekannter Lösung sein. Durch Vergleich von Systemlösung und korrekter Fallösung kann so eine Aussage über die Qualität des Problemlöseverhaltens gemacht werden.

Eine weitere Methode, die unabhängig von einer Studie durchgeführt werden kann, ist die Verwendung von *Kriterienkatalogen* zur Bewertung von CBT–Systemen. Viele Gruppen innerhalb der CBT–Forschung haben versucht, gemeinsam Kriterien aufzustellen, die ein gutes Programm auszeichnen. Oft sind es aber auch Kriterien, die weniger der Bewertung sondern hauptsächlich der Einordnung der Programme dienen, so daß der Lernen-

de anhand dieser sehen kann, welches System seinen Anforderungen am besten entspricht.

Bekannte Kriterienkataloge in diesem Bereich sind:

- Qualitätskriterien für elektronische Publikationen in der Medizin der GMDS[1] CBT–Arbeitsgruppe,
- Kriterienkatalog für CBT–Programme der AKAB[2],
- IEEE[3] Standards für Tutorsysteme.

Als letztes soll hier die in [MARK and GREER 1993] erwähnte *Sensitivitätsanalyse* genannt sein, die sich darauf stützt, daß der Vorteil der computerunterstützten Lehre in der individualisierten Anpassung an den Lernenden liegt. Hier wird untersucht, in wie weit das System auf verschiedene Benutzereingaben unterschiedlich reagiert. Viele Systeme bezeichnen dieses Verhalten als Adaptivität, wobei meistens nur die Oberfläche und nicht das interne Verhalten angepaßt wird.

4.1.2.2 Studienabhängige Methoden

Das primäre Ziel aller Lehrprogramme ist es, den Wissenszuwachs bei den Lernenden zu fördern. Deshalb sind Studien, in denen Lernende mit dem CBT–Programm arbeiten, und der Vergleich des Wissensstands vor und nach der Arbeit mit dem System mit die wichtigsten Methoden bei der Evaluation solcher Systeme. Aber nicht nur der reine Wissenszuwachs, sondern auch andere Meßwerte wie beispielsweise die Motivation der Lernenden sind bei diesen Studien interessant und können zu einem verbesserten System in der nächsten Version führen.

Bei der Durchführung von Pilotstudien können verschiedene Methoden eingesetzt werden, um die Interaktion von Benutzer und System zu untersuchen. Bei der *Verhaltensbeobachtung* werden die Lernenden durch die Studienleiter bei dem Umgang mit den System beobachtet. Die Beobachter betrachten verschiedene Faktoren, die sie anschließend auswerten können. Mit dieser Methode kann man besondere Unsicherheiten oder Schwachpunkte in der Benutzungsoberfläche herausfinden. Ein weiterer Schritt in dieser Richtung ist das *Monitoring*, bei dem alle Interaktionen des Lernenden durch das System aufgezeichnet werden. Diese Verlaufsprotokolle können später bezüglich bestimmter Fragestellung genau ausgewertet werden. Beispielsweise kann so die Bearbeitungszeit mit den angeforderten Entscheidungshilfen in Beziehung gesetzt werden.

[1] Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V.
[2] AKAB = Arbeitskreis der Automobil Branche
[3] IEEE = Institute for Electrical and Electronics Engineers

Durch Verhaltensbeobachtung und Monitoring können die äußeren Aktivitäten des Lernenden gut untersucht werden. Oft interessieren aber die internen gedanklichen Vorgänge bei dem Lernenden, die so nicht erfaßt werden können. Eine Möglichkeit, Einblick in die internen Problemlösevorgänge zu bekommen, sind *Think–aloud–Protokolle*. Die Lernenden werden aufgefordert, bei der Arbeit mit dem System laut zu denken und so ihre Probleme, Strategien und Hypothesen zu formulieren. Die aufgezeichneten Protokolle können später entsprechend ausgewertet werden.

Neben diesen fallbegleitenden Methoden gibt es andere Möglichkeiten, die Akzeptanz und Probleme bei der Arbeit mit dem System zu untersuchen, die nach Abschluß der Fallbearbeitung angewandt werden können. In *Interviews* mit den Lernenden können Probleme individuell aufgespürt und analysiert werden. Für eine gute Durchführung sollte den Interviewern ein Leitfaden zur Verfügung gestellt werden, an den sie sich halten sollen, bis sie bei Problemen detaillierter auf die Befragten eingehen. Bei dieser Art der Evaluation kommt es besonders auf die kommunikativen Fähigkeiten bei Interviewer und Lerner an. Kann der Lerner seine Probleme nicht formulieren oder entstehen Probleme bei der menschlichen Kommunikation zwischen den beiden Partnern, kann dies das Ergebnis der Untersuchung beeinflussen. Dieses Problem besteht aber bei allen Evaluationsformen, die auf menschlicher Kommunikation bzw. Einschätzung beruhen, da hier die grundsätzlichen Einschränkungen der Subjektivität bestehen bleiben. *Fragebögen*, in denen die Lernenden ihre Erfahrungen und Bewertungen eintragen können, vermeiden zumindestens die Beeinflussung des Lernenden durch Sympathie oder Apathie gegenüber dem Interviewer. Allerdings können sie hier ihre persönliche Meinung nur durch Auswahl aus bestimmten Alternativen und evtl. einem Kommentar formulieren. Diese Einschränkung hat aber auch den Vorteil, daß sie für den Studienteilnehmer weniger zeitintensiv ist, was vor allem bei realen Einsätzen ein besonderer Pluspunkt ist.

Objektive Bewertungen für den Erfolg einer Lernumgebung stellen die *Leistungsmessung* dar. Die Studenten legen vor und nach der Studie einen Leistungstest ab und der Vergleich der Ergebnisse stellt den Wissenszuwachs dar. Bei einer vergleichenden Evaluation können die Ergebnisse zusätzlich mit denen der Kontrollgruppe in Relation gesetzt werden. Die Kontrollgruppe wird in der Studie mit konventionellen Lehrmethoden, anderen Systemen oder einer anderen Variante des untersuchten Systems unterrichtet. Wichtig bei diesem Vergleich sind aber auch die Lernbedingungen, wie beispielsweise die vorhandene Lernzeit, die bei beiden Gruppen ungefähr gleich sein sollten. Deshalb sollte die Gruppeneinteilung zufällig und die Gruppenstärke so groß sein, daß die Gruppe als repräsentativ angesehen werden kann. Da diese Voraussetzungen in realen Einsätzen meist nicht zu erfüllen sind, besteht die Möglichkeit, einen Crossover–Versuchsplan zu

benutzen, wie beispielsweise in [AUHUBER 1998]. Hier werden zwei Gruppen gebildet, die abwechselnd in zwei Wissensgebieten einmal alternativ und einmal mit der zu untersuchenden Lehrmethodik unterrichtet werden. Eine andere Möglichkeit ist auch, den Wissenszuwachs der Studenten an ihren Leistungen in anderen ähnlichen Wissensgebieten zu vergleichen, die auf alternative Weise unterrichtet wurden.

Der Erfolg einer Feldstudie hängt von vielen Details ab und sollte deshalb sorgfältig vorbereitet werden. [SHUTE and REGIAN 1993] nennen dazu verschiedene Schritte bei der Planung solcher Evaluationen, die Probleme vermeiden sollen:

1. Bestimme die Ziele des Systems: Es soll eine klare Aussage getroffen werden, was das System vermitteln möchte und wie es das tun will.

2. Bestimme die Ziele der Studie: Hier soll eine klare Aussage darüber gemacht werden, warum diese Studie gemacht wird und wozu das Ergebnis verwendet werden soll.

3. Suche das Evaluationsdesign aus, das den beiden Zielen angemessen ist.

4. Instantiiere das Studiendesign mit den angemessenen Meßmethoden, Anzahl und Typ von Lernenden und Kontrollkonditionen.

5. Bereite die Durchführung der Studie sorgfältig und logisch vor: Die Auswahl der betreuenden Tutoren und deren Schulung fallen genauso unter diesen Punkt wie die generelle Organisation (z. B. Raumsuche, Verpflegung), die den Verlauf einer Studie später stören könnten (z. B. Doppelbelegungen von Räumen).

6. Mache einen Pilottest der Studie: Bei einem kleinen Test mit ein oder zwei Personen, die das Studiendesign ausprobieren, können schon Effekte auftreten, die in der späteren Studie zu echten Probleme führen könnten.

7. Lege die primäre Datenanalyse während der Studienplanung fest: Da ja bereits Ziele der Studie festgelegt wurden, kann auch eine klare Aussage getroffen werden, welche Daten hierfür notwendig sind. Durch diese Überlegung kann vermieden werden, daß evtl. wichtige Daten nicht erhoben werden. Durch die Festlegung der Datenanalyse wird auch sichergestellt, daß die Daten in verarbeitbarer Form erhoben werden.

Die sorgfältige Planung einer Studie und die Entscheidung für die angemessene Form und entsprechende Methoden sind nicht nur wichtig für die

problemlose Durchführung einer Studie, sondern auch für die Zuverlässigkeit und Aussagekraft der gewonnenen Daten. Bei Studien in realen Einsätzen können allerdings oft nicht alle Voraussetzungen optimal erfüllt werden, die in einer Laborevaluation möglich gewesen wären. Trotzdem sind die Aussagen, die in einer realen Umgebung gefunden werden, oft interessanter, wenn auch insgesamt mehr Spielraum für Spekulationen bleibt.

4.2 Evaluation der Entwicklungsumgebung

Die in dem vorherigen Kapitel genannten Evaluationsmethoden bezogen sich auf CBT–Systeme und weniger auf die Werkzeuge, mit denen man solche Systeme erstellt. Alle Evaluationsformen und einige Evaluationsmethoden sind aber auf die Werkzeuge übertragbar. Meistens werden die Werkzeuge durch die Evaluation der entstandenen Systeme bewertet. Natürlich ist das auch durch den Anspruch der Systeme gerechtfertigt, daß man mit ihrer Hilfe gute tutorielle Programme entwickeln kann. Trotzdem sollte in Zukunft auch die Bewertung der Autorensysteme stärker beachtet werden.

Das Autorenwerkzeug D3TRAINER stellt aber wieder eine Sonderform im Gesamtrahmen der ITS–Autorensysteme dar. Durch den Aufbau auf den Expertensystemshellbaukasten D3 ist der D3TRAINER kein Autorenwerkzeug im eigentlichen Sinn, sondern benutzt die bereits erstellten Wissens– und Fallbasen zur Generierung fallbasierter Trainingssysteme. In diesem Fall müßte also eine Evaluierung der Wissenserwerbskomponente von D3 betrachtet werden, die aber schon in [GAPPA 1995] abgehandelt wurde. Um trotzdem eine Aussage über die Qualität des D3TRAINERS als Autorenwerkzeug machen zu können, wird dieser im folgenden mit Hilfe der in Abschnitt **??** aufgestellten Kriterien eingeordnet und mit den dort vorgestellten ITS–Autorensystemen verglichen.

Anwendungsbreite und Einordnung der entstehenden Systeme: Die durch den D3TRAINER entstehenden Systeme sind fallbasierte Trainingssysteme, die den problemlöseorientierten Systemen zuzuordnen sind. Alle diagnostischen Probleme können behandelt werden, wobei der Schwerpunkt der Systeme auf der diagnostischen Problemlösung und deren Kritik liegt. Lehrbuchwissen kann über Anbindung an Texte, Bilder oder elektronische Lehrbücher vermittelt werden, wird aber nur als Zusatzoption angeboten.

Führung und Unterstützung der Wissensakquisition: Der Autor der entstehenden Systeme ist ein Fachbereichsexperte ohne Programmierkenntnisse, der sein Problemlösewissen, eine Auswahl an Fällen und gegebenenfalls Bildsysteme zu den Fällen eingibt. Diese Eingabe

wird immer durch graphische Editoren geführt, wobei der Experte durch verschiedene Funktionalitäten, wie beispielsweise automatisches Testen mit Fällen, unterstützt wird.

Studentenmodellierungsmethoden: Eine Modellierung des Studentenwissens findet in dem D3TRAINER grundsätzlich nicht statt. Durch die Variabilität der Systemparameter können jedoch durch den Autor gewissen Stereotypen angegeben werden.

Lehrstrategien: Auf die verschiedenen Lehrstrategien wird hier nicht näher eingegangen, sondern auf Abschnitt **??** verwiesen, in dem die multiplen Lehrstrategien genauer diskutiert wurden.

Entwicklungsdauer, Wiederverwendbarkeit und Wartung: Die Erweiterbarkeit einer D3–Wissensbasis ist sehr einfach und ohne starke Beeinflussung der restlichen Wissenselemente möglich. Die Eingabe eines neuen Falles hängt nur von der Anzahl der einzugebenen Symptome ab und liegt durchschnittlich zwischen 10 und 20 Minuten. Zum Erstellen eines Bildsystems bei vorhandenen Bildmaterial benötigt man etwa 30 Minuten. Für den Aufbau einer Wissensbasis können solche zeitliche Aussagen nur schwer getroffen werden, da sie sehr stark von Größe und Einarbeitungszeit des Experten abhängen. Unter günstigen Bedingungen dauert der Aufbau einer Wissensbasis abhängig von Umfang und Komplexität der Domäne Wochen bis Monate.

Das hinter dem D3TRAINER stehende Konzept richtet sich also mehr nach dem ‚special purpose'–Ansatz und nicht nach dem der universell einsetzbaren Systeme, wie beispielsweise Redeem oder EON. Durch die Konzentration auf die diagnostische Problemlösung kann der Autor in dem Entwicklungsprozeß sehr stark unterstützt werden und auch die grundsätzlich festgelegte didaktische Strategie kann soweit sinnvoll durch den Experten variiert werden. Im Gegensatz zu den anderen genannten Autorenwerkzeugen mit internem Modell besitzt der D3TRAINER echtes Problemlösewissen, was besonders bei der automatischen Erklärungsfähigkeit enorme Vorteile mit sich bringt.

4.3 Evaluation des Werkzeugs

Neben der Beschreibung des D3TRAINERS als Entwicklungsumgebung in dem letzten Abschnitt, ist es natürlich auch wichtig, wie er als Werkzeug zu bewerten ist. Als Grundlage für diese Evaluation wurden fünf Autoren medizinischer Anwendungen mittels Fragebogen um ihre Meinung gebeten. Die Befragung betraf die drei Hauptpunkte *Benutzeroberfläche, didaktische Aspekte* und *Feedback-Algorithmen*. Neben der konkreten Bewertung

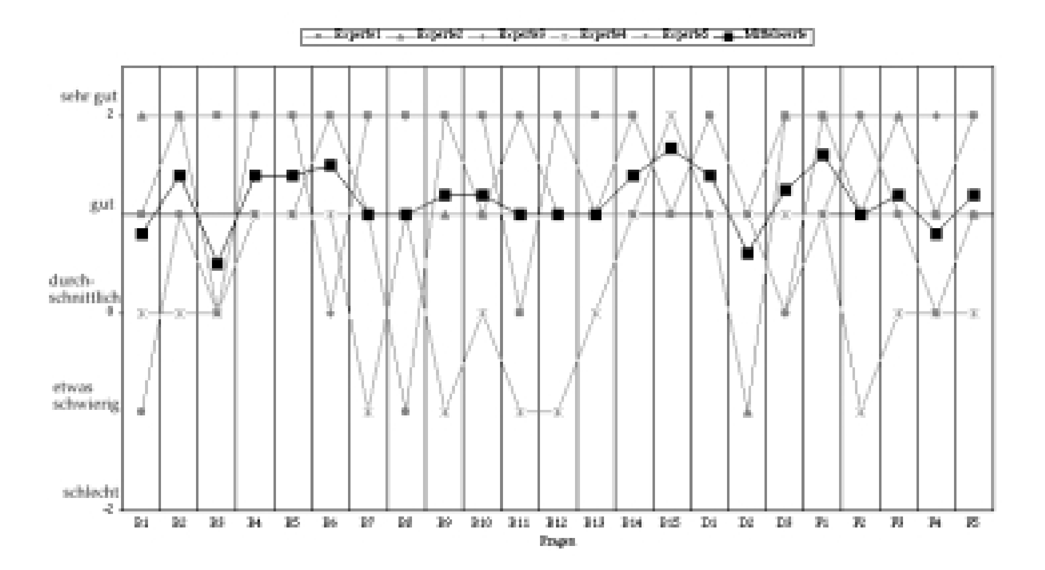


Abbildung 4.2: Bewertungen und Mittelwert des Fragebogens zur Bewertung des D3TRAINERS als Werkzeug.

einzelner Aspekte des D3TRAINERS durch die Kategorien *schlecht* (-2), *etwas schwierig* (-1), *durchschnittlich* (0), *gut* (1) und *sehr gut* (2), konnte zu allen Fragen noch ein freier Kommentar gegeben werden. Die genauen Werte sind der Abbildung 4.2 zu entnehmen, in der auch die Mittelwerte eingezeichnet sind.

Der erste Fragenkomplex betraf die Benutzeroberfläche des D3TRAINERS und bestand aus folgenden Einzelfragen (hinter den Fragen stehen die Abkürzungen aus der Abbildung):

- Wie bewerten Sie die intuitive Benutzbarkeit der Traineroberfläche? (B1)
- Wie bewerten Sie die Umsetzung der Symptompräsentation des Falles, besonders aber:
  - Benutzung der Klapphierarchie? (B2)
  - Schematische Symptom=Wert–Darstellung? (B3)
  - Farbgebung für normale / abnorme Symptome? (B4)
- Wie fanden Sie konkret die Gestaltung folgender Fenster:
  - Diagnosenauswahl? (B5)
  - Testauswahl? (B6)
  - Eingabe der Symptomerkennung? (B7)

  - Feedbackfenster zur Verdachtsdiagnosen? (B8)
  - Feedbackfenster zur Begründung einer Verdachtsdiagnose? (B9)
  - Feedbackfenster zur Testauswahl? (B10)
  - Feedbackfenster zur Begründung der Testauswahl? (B11)
  - Feedbackfenster zur Symptomerkennung? (B12)
  - Feedbackfenster zur Gesamtbewertung? (B13)

- Wie fanden Sie den Einsatz von Navigationsmechanismen (Buttons, Menüs etc.) im D3TRAINER? (B14)

- Wie bewerten Sie die Einbettung der Kostenbetrachtung (Kosten, Risiko und Zeitaufwand)? (B15)

Durchschnittlich wurden alle Fragen zur Benutzeroberfläche als gut bezeichnet, nur bei den Fragen zur intuitiven Benutzbarkeit und der Symptom=Wert–Darstellung gab es leichte Tendenzen zu der Bewertung *durchschnittlich*. Vereinzelt schlechte Bewertung wurden für die Gestaltung mancher Feedbackfenster vergeben. Überdurchschnittlich gut waren die Bewertungen zu der Farbgebung der Symptome, der Gestaltung der Diagnosenauswahl und zur Integration der Kostenbetrachtung. Bei den Kommentaren wurde durchgängig eine Online–Hilfe gewünscht, sowie einzelne graphische Erweiterungen (z. B. Darstellung der Laborwerte als Tabelle) angeregt. Mehrere Fachexperten wünschten sich eine an das Themengebiet angepaßte Oberfläche, so daß anwendungsspezifische Symbole in die Oberfläche eingebunden werden. In einer medizinischen Anwendung wären das beispielsweise Apparaturen, weiße Kittel oder Stethoskope. Einer der Befragten formulierte auch den Wunsch seiner Anwender nach einem Nebeneinander von Symptom-, Diagnose und Therapieebene.

Die Fragen für den Bereich *Didaktische Aspekte* setzten sich folgendermassen zusammen:

- Sind die vorhandenen Mechanismen zur Symptomanforderung (ganzer Fall, Aufwandsgruppen, Frageklassen, Einzelanforderung) ausreichend? Haben Sie zusätzliche Wünsche? (D1)

- Wie bewerten Sie die didaktische Umsetzung der realen Problemlösesituation (Klinikalltag) im D3TRAINER ? (D2)

- Wie bewerten Sie die Möglichkeit, informelles Lehrbuchwissen in die Problemlöseumgebung zu integrieren? (D3)

Auch in dem Bereich der didaktischen Aspekte lagen die Bewertungen im Bereich *gut*. Einzige Ausnahme war die Bewertung eines einzelnen Expertens für die Umsetzung der realen Problemlösesituation, die von ihm

als *etwas schwierig* bezeichnet wurde. Seine Begründung dafür war, daß der therapeutische bzw. interventionelle Aspekt vergleichsweise wenig berücksichtigt sei. Auffallend bei der Bewertung war, daß keine zusätzlichen Mechanismen zur Symptomanforderung gewünscht wurden (D1), so daß davon ausgegangen werden kann, daß die bereitgestellten Möglichkeiten das benötigte Spektrum abdecken.

Bezüglich der *Feedback–Algorithmen* wurden folgende Fragen gestellt: Wie bewerten Sie das generierte Feedback zur Bewertung der

- Auswahl von Verdachtsdiagnosen? (F1)
- Begründung einer Verdachtsdiagnose? (F2)
- Auswahl von Tests? (F3)
- Begründung der ausgewählten Tests? (F4)
- Symptomerkennung? (F5)

Die Bewertung der Feedback–Algorithmen, in denen ja die eigentliche Intelligenz des D3TRAINERS liegt, wurde ebenfalls als *gut* bewertet. Überaschend war hier nur, daß der Experte, der im Verlauf der Entwicklung der Algorithmen am meisten Einfluß nehmen konnte, schlechtere Bewertungen vergab als die Experten, die diese Algorithmen nur nutzten. Eine Bewertung als *etwas schwierig* wurde allerdings nur für den Algorithmus für das Feedback zur Begründung von ausgewählten Verdachtsdiagnosen vergeben. Der Kommentar zu dieser Einschätzung betrifft die Tatsache, daß als Grundlage für das Feedback nur heuristische Regeln verwendet werden können. Durch eine Einstellung können aber bereits in der vorliegenden Version des D3TRAINERS auch die geforderten überdeckende Regeln benutzt werden. Dies gilt jedoch nicht für die Bewertung der Auswahl von Therapien.

In den Kommentaren wurden auch die Probleme bei der Testauswahl angesprochen, die durch die intern vorgegebene Reihenfolgebestimmung des Expertensystems entstehen. Der Algorithmus wurde hier zwar schon verändert, eine optimale Lösung scheint aber noch nicht gefunden. Vereinzelt wurde auch eine stärkere Modifizierbarkeit der Algorithmen allgemein gewünscht.

Zum Abschluß des Fragebogens wurden den Befragten noch die Möglichkeit gegeben, sich zu den Stärken und Schwächen des D3TRAINERS zu äußern und weitere Kommentare zu formulieren. Als Stärken wurde das Training von rationellem Vorgehen nach Kosten–Nutzen–Gesichtspunkten genannt, sowie die Tatsache, daß jeder aktuelle Patient ohne großen Aufwand als Tutorfall erfaßbar ist. Hierbei wurde die Vielzahl der trainier-

baren Diagnostikentscheidungen bei Offenhalten des gesamten Alternativenspektrums (keine Einengung auf wenige MC–Alternativen bei den Studentenaktionen) gelobt. Auch die Einbeziehung von Lehrbuchtexten und die Möglichkeit, Tutandenaktionen durch das System kritisieren zu lassen, wurde hervorgehoben. Der D3TRAINER führe zu schlußfolgernden Denken hin und diszipliniert diagnostische Entscheidungen. Die Tatsache, daß das Gesamtsystem D3 nicht nur als Trainingssystem sondern auch zur Diagnoseunterstützung, Arztbriefgenerierung etc. einsetzbar ist, wurde ebenfalls genannt.

Die größten Probleme sehen die Anwender in dem Aufwand zur Erstellung einer aussagekräftigen Wissensbasis, sowie der Unterrepräsentation von Therapien und Zeitverläufen. Problematisch bei der Fallbearbeitung durch den Studenten sei auch, die Übersicht zu behalten, wenn zu viele Fenster gleichzeitig offen sind.

Als zusätzlicher Kommentar wurde nur genannt, daß Änderungswünsche einzelner Anwender individuell eingebaut werden können, da nicht jede Änderung des Systems gleich gut für alle Applikationen ist.

Insgesamt zeigte die Befragung, daß die Umsetzung des fallbasierten Lernens in der Diagnostik im D3TRAINER durchgängig gut gelungen ist, aber noch Verbesserungsmöglichkeiten bei den einzelnen Aspekten bestehen. Der schwierigste Teil, die generierte Bewertung der Studentenaktionen, wurde ebenfalls als gut bewertet. Die Anwender zeigten sich also zufrieden mit dem System und seinen gesamten Möglichkeiten.

4.4 Evaluationen der Anwendungen

Die mittels des D3TRAINERS generierten Anwendungen wurden in verschiedenen Umgebungen angewandt bzw. evaluiert. Federführend bei den Evaluationen waren dabei meist die Wissensbasisentwickler, weshalb es keine einheitliche Evaluierung aller Anwendungen gibt. Trotzdem kann durch Vergleich der einzelnen Evaluationen eine Gesamtaussage über den D3TRAINER als Werkzeug gemacht werden. Im folgenden werden die Evaluationen der damit entwickelten fallbasierten Trainingssysteme vorgestellt.

4.4.1 Rheumatologie

Der *RheumaTutor*, der von Prof. Schewe an der Poliklinik der Universität München entwickelt wird, wurde in verschiedenen Szenarien evaluiert. Das rheumatologische System ist in mehreren Varianten und in verschiedenen Versionen des D3TRAINERS angewandt worden. In der Wissensbasis sind etwa 100 rheumatologische Krankheiten und 40 Therapi-

en abgedeckt. Die Wissensbasis wurde in verschiedenen Studien auf ihre Problemlösefähigkeit überprüft (siehe [GOOS and SCHEWE 1993], [SCHEWE and SCHREIBER 1993]).

Seit dem Wintersemester 1994/95 verwendet Prof. Schewe den *RheumaTutor* in einem Pflichtkurs über innere Medizin. Im Sommer 1998 wurde das *RheumaSeminar*, eine Variante des *RheumaTutor*, zum ersten Mal in einem prototypischen Versuch für die Fortbildung niedergelassener Ärzte (Qualitätszirkel) verwendet. Der *RheumaTutor* mit 25 multimedialen Fällen wird inzwischen durch den Ullstein Medical Verlag kommerziell vertrieben. Mehrere Gutachten anhand des GMDS–Kriterienkataloges erstellt und miteinander verglichen. Die Evaluationen dieser Einsätze und ihre Ergebnisse werden im folgenden genauer diskutiert.

4.4.1.1 Poliklinik München

Die Einbindung des *RheumaTutor* in dem Kurs über innere Medizin an der Poliklinik der Universität kann wie folgt beschrieben werden: Während des Semesters durchlaufen die Studenten in Kleingruppen (4 Personen) mehrere Stationen der inneren Medizin, wie beispielsweise die Rheumatologie, die Kardiologie oder die Hepatologie. Dabei sind sie für zwei Wochen jeweils an zwei Tagen in einer Station.

In der Rheumatologie werden sie am ersten Tag mit einem realen Patienten konfrontiert, an dem sie zusammen mit Prof. Schewe die Anamnese und einige körperliche Untersuchungen durchführen. Dabei wird, entsprechend den Prinzipien des problemorientierten Unterrichts, der Fall durch die Studenten in der Diskussion mit dem Dozenten (Tutor) besprochen. Nachdem die Studenten die Rheumaeinheit verlassen haben, wird der Patient vollständig als Fall in das Expertensystem eingegeben. Mit Einverständnis des Patienten werden Bilder (Photographien krankhafter Körperregionen) gemacht und ein entsprechendes Bildsystem zu dem Fall erstellt.

Aus der umfangreichen rheumatologischen Fallsammlung (ca. 1000 klinische Fälle) wird mittels der fallbasierten Komponente des D3–Expertensystems ein Fall mit ähnlicher Symptomatik aber einer anderen Diagnose (Differentialdiagnose) gesucht. Beide Fälle werden den Studenten am nächsten Tag in dem *RheumaTutor* präsentiert. In der Kleingruppe bearbeiten die Studenten nun gemeinsam die beiden Fälle.

Seit dem Sommersemester 1997 steht zusätzlich jedem Studenten eine CD mit 25 rheumatologischen Fällen zur Verfügung, die sie auf ihrem Heimcomputer[4] bearbeiten können. Die an der Poliklinik München praktizierte Lehre wurde sehr stark nach den problemorientierten Prinzipien ausgerichtet. Sowohl die Arbeit an realen klinischen Fällen, der persönliche Kon-

[4]Betriebssystem Windows95 oder höher mit CD-Rom Laufwerk.

Semester	Anzahl der Studenten
Wintersemester 1994 / 1995	22
Sommersemester 1995	19
Sommersemester 1997	27
Sommersemester 1998	14

Tabelle 4.1: Anzahl der teilnehmenden Studenten an den Studien nach Semestern.

takt mit den Patienten als auch die Arbeit in der Gruppe unterstützt die Ausbildung sozialer Fähigkeiten bei den Studenten und verstärkt deren Motivation zum Lösen der Fälle.

Vor und nach dem gesamten Kurs wurden Leistungstests erhoben und verschiedene Fragebögen ausgefüllt, die in der Evaluationsstudie ausgewertet werden. Eine Abspaltung einer Kontrollgruppe ist in diesem Szenario aus mehreren Gründen nicht möglich[5]. Um trotzdem einen Vergleich der studentischen Leistungen in einem nicht computerunterstützten Kurs durchführen zu können, werden die Leistungen der gleichen Studenten in einem anderen Gebiet der inneren Medizin (z. B. Kardiologie) herangezogen. Den Untersuchenden war aber bewußt, daß das Lernverhalten in unterschiedlichen Fachgebieten völlig unterschiedlich sein kann und deshalb die Ergebnisse mit der erforderlichen Vorsicht interpretiert werden müssen. Der genaue Verlauf der Untersuchung wird in Abb. 4.3 nochmals verdeutlicht.

Die Untersuchungen an der Poliklinik können als formative Makroevaluationen im realen Einsatz betrachtet werden. Die Leistungen der Studenten werden mit anderen Gebieten der inneren Medizin, die ohne Computerunterstützung unterrichtet wurden, verglichen. Da die Studie durch den Wissensbasisentwickler organisiert wird, liegt eine innere Evaluation vor. Eine Ausnahme bildet das Sommersemester 1997, da hier die Studie im Rahmen einer Diplomarbeit der Psychologie [LANGER 1997] durchgeführt wurde. Inzwischen wurden auch im Wintersemester 1998 / 1999 und im Sommersemester 1999 weitere Kurse mit dem *RheumaTutor* durchgeführt, bei denen aber keine Fragebögen erfaßt wurden.

Die Anzahl der teilnehmenden Studenten verteilte sich in den Semestern wie in Tabelle 4.1 dargestellt. Für eine genaue Beschreibung der Fragebögen und numerischen Ergebnisse sei auf die Veröffentlichungen (siehe [SCHEWE et al. 1996], [LANGER 1997], [PUPPE et al. 1998], [SCHEWE et al. 1999]) hingewiesen, da in diesem Rahmen nur die Gesamtergebnisse zusammengefaßt werden.

Die Fragebögen der Studie waren so aufgeteilt, daß die Studenten Fragen zu dem *RheumaTutor* und zum generellen Einsatz computerunterstützter Systeme in der Ausbildung bewerten sollten. Hierfür wurde eine nume-

[5] Hier ist grundsätzlich der doppelte Vorbereitungsaufwand bei nur einem Dozenten und das Problem der Diskriminierung zu nennen.

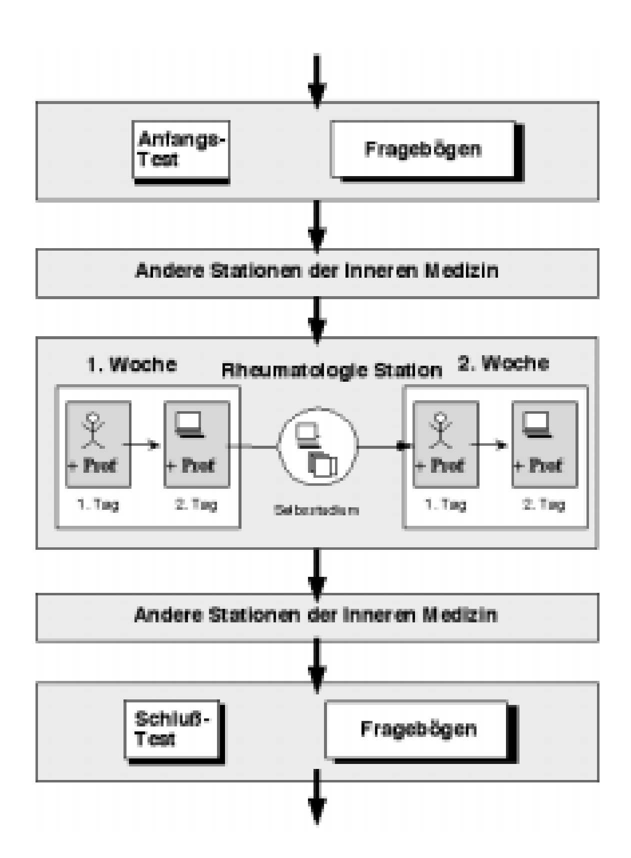


Abbildung 4.3: Durchführung der Untersuchung an der Poliklinik München: Die Studenten füllen zu Begin des Semesters einen Wissenstest und einen Fragebogen aus. Während des Semesters durchlaufen sie alle Stationen der Inneren Medizin, darunter auch die Rheumatologie Station. In jeder Station sind sie zwei Wochen lang jeweils zwei Tage zum Unterricht. In der Rheumatologie werden sie am ersten Tag mit einem realen Patienten konfrontiert, den sie gemeinsam mit dem Professor untersuchen. Am nächsten Tag bekommen sie diesen und einen ähnlichen Patientenfall in dem Computerprogramm *RheumaTutor* präsentiert und müssen diesen bearbeiten. Dabei ist der Professor anwesend und greift in Fallbesprechung und Diskussion der Studenten steuernd ein. In der zweiten Woche wiederholt sich diese Vorgehensweise, wobei sich der Professor am zweiten Tag weniger in die Diskussion einbringt. Zwischen der ersten und zweiten Woche bleibt den Studenten Zeit für das Selbststudium zu Hause mit Literatur und ab Sommersemester 1997 auch mit dem *RheumaTutor* am Computer zu Hause.

rische Einschätzung zwischen 0 (schlecht) und 10 (optimal) gewählt. In Abb. 4.4 sind einige Daten veranschaulicht, die genauen Daten sind im Anhang 6.1 dargestellt.

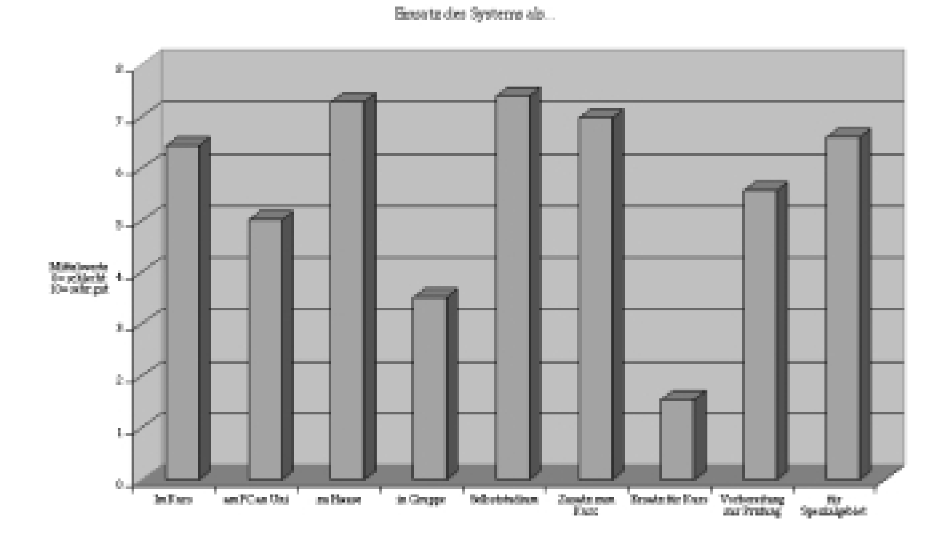


Abbildung 4.4: Einschätzung der Studenten für die optimale Einsatzumgebung des D3TRAINERS bei der medizinischen Ausbildung.

Deutlich überdurchschnittlich gute Beurteilungen wurden in folgenden Fragen gegeben:

- der Einsatz eines Computerlernprogramms ist sinnvoll,
- das vorgestellte Programm ist für den dargestellten Einsatz nahezu optimal,
- die vorgestellte Nutzung des Systems im Kurs an der Universität wird befürwortet,
- die Nutzung des Programms für das Selbststudium zuhause ist am sinnvollsten,
- das Programm als Ergänzung zum Kurs ist vernünftig,
- es eignet sich eher zum Erlernen von Spezialgebieten,
- es motiviert zum Erlernen unbekannter Gebiete,
- es ist als Lernhilfe sinnvoll,
- es ist nicht langweilig,
- es dient mehr der Fallanalyse,
- es übt das Erlernen der wichtigsten Anamnesefragen,
- es schult mehr diagnostische Fähigkeiten,

- es dient der Unterstützung zur richtigen Diagnose und
- es regt zum selbständigen Denken an.

In folgenden Bereichen gaben die Studenten mittlere Bewertungen:

- Textverständlichkeit,
- Textgliederung,
- Augenfreundlichkeit der Texte,
- Erlernbarkeit der Software,
- Anregung mit dem Programm sich weiter zu beschäftigen,
- Nachvollziehbarkeit der Computerentscheidungen,
- Freude bei der Arbeit und
- Einschätzung der Nützlichkeit für das Studium.

Negativ wurden beurteilt:

- technische Probleme.

Die einzelnen Semester unterschieden sich nur marginal voneinander, ebenso wenig scheint das Geschlecht oder das Vorhandensein eines Computers zu Hause einen Einfluß auf die Bewertung zu haben. Studenten, die deutlichere Probleme bei der Benutzung der CD am Heimcomputer hatten, beurteilten generell Fragen schlechter, besonders aber bezüglich des *RheumaTrainers* und der Anwendung von Lernsoftware in der Ausbildung allgemein.

Eine Zunahme des Wissens wurde sowohl für das Wintersemester 1994/95 als auch für das Sommersemester 1998 beobachtet. Bei zwei unterschiedlichen Fällen mit einer vermutlichen *Kollagenose* wurden am Ende des Kurses deutlich häufiger richtige Angaben zur Anamnese, klinischen Untersuchung, Laboruntersuchungen und bildgebenden Verfahren beobachtet als am Anfang des Semesters. Ebenso zeigten sich deutlich häufiger richtige Antworten bei einer Abklärung von zwei Fällen mit einer *Monarthritis*, es konnten jedoch keine Unterschiede in der Abklärung einer *Gicht* nachgewiesen werden.

Der *RheumaTrainer* wird von den Studenten überwiegend positiv beurteilt, insbesondere wenn er wie in dieser Studie als integrativer Bestandteil der Ausbildung eingesetzt wird. Studenten bevorzugen verständlicherweise den realen Patienten deutlich gegenüber der Arbeit mit einem Computerprogramm, weshalb die kombinierte Anwendung eines Lernprogramms

als Ergänzung zur üblichen Patientenvorstellung besonders gut akzeptiert wird. Die Studenten wurden nicht ausgewählt, weil sie eine besondere Neigung zum Lernen mit Computern hätten, sondern sie wurden durch das Dekanat der Klinik zur Ausbildung zugewiesen. Ein Teil der Studenten lehnt aber den Computer als neues Lernmedium ab. Ein wesentlicher Einflußfaktor dafür waren die technischen Probleme bei der Handhabung des Programms. Zum Teil lagen die Probleme daran, daß die Studenten Computer zur Verfügung hatten, die einen zu kleinen Arbeitsspeicher aufwiesen, zum Teil auch an der fehlenden schriftlichen Einführung der Studenten in die Handhabung des Programms. Diese Probleme konnten in der Zwischenzeit behoben werden.

Der meßbare Lernerfolg sowohl unmittelbar im Anschluß an den Kurs als auch bei der Abschlußprüfung (teilweise Monate nach dem Kurs) spricht dafür, daß mit dem Kurs Lerninhalte vermittelt wurden, die über längere Zeit erkenn- und meßbar waren. Dies trifft insbesondere in Bereichen zu, in denen die Studenten am Beginn kaum Kenntnisse vorwiesen (*Kollagenosen*), weniger in Bereichen, wo auch schon zu Anfang des Kurses richtige Ideen vorlagen (*Gicht*).

4.4.1.2 Qualitätszirkel

Im Rahmen einer Initiative zur Verbesserung der ärztlichen Fortbildung wurde in Zusammenarbeit mit Prof. Rienhoff (Universität Göttingen), dem Verlag Ullstein Medical und verschiedenen Ärztekammern ein prototypischer Einsatz in Qualitätszirkeln geplant und durchgeführt [RIENHOFF 1998]. In diesen Veranstaltungen wird für niedergelassene Ärzte anhand von Fallbeispielen ein Spezialgebiet der Medizin wiederholt. Bisher wurde für diese Veranstaltungen kein Computerprogramm benutzt, vielmehr wurde mit Dias und Fallbeschreibungen gearbeitet.

In dem Konzept für diese Qualitätszirkel werden zwei Treffen geplant: Bei dem ersten Treffen wird ein Fall mit den Ärzten gemeinsam am Computer bearbeitet. Wie auch bei dem Einsatz mit den Studenten bietet das Programm die Basis für Diskussionen und Austausch zwischen den Ärzten. Für diesen Einsatz wurden vier Fälle speziell ausgesucht und in einer anderen didaktischen Variante angeboten. Dieses *Rheumaseminar* wird den Ärzten als CD mitgegeben und sie werden gebeten, die drei verbleibenden Fälle bis zum nächsten Treffen zu bearbeiten. Bei dem zweiten Treffen bekommen sie dann eine weitere CD mit vier neuen Fällen. Weiterhin werden alle Benutzeraktionen protokolliert und abgespeichert (Fallverlaufsprotokoll), so daß die Teilnehmer die Diskette mit den Daten an die Seminarleiter zur Endauswertung senden können.

Hier kann von einer summativen Makroevaluation ausgegangen werden, die mittels Fragebögen und Monitoring durchgeführt wird. Ein Vergleich

Wo?	Personen 1. Treffen	Personen 2. Treffen	Anzahl Frage-bögen	Anzahl Fallproto-kolle
Ärzteverein Holzminden	13	11	7	66
Krankenhaus Northeim	13	11	1	44

Tabelle 4.2: Genaue Daten der Qualitätszirkeltreffen.

findet nicht statt. Da die Auswertung durch Mitarbeiter der Universität Göttingen durchgeführt wird, kann von einer äußeren Evaluation gesprochen werden.

In dieser Evaluationsstudie wurden zwei Qualitätszirkel mit jeweils zwei Treffen durchgeführt, deren genaue formale Daten in Tabelle 4.2 beschrieben sind.

Die Bewertungen in den Fragebögen konnten zwischen -2 und +2 gegeben werden, wobei ein hoher Wert immer eine positive Aussage war. Beispielsweise sollten die Befragten einschätzen, ob sie beim nächsten Patienten mit entsprechender Symptomatik anders vorgehen würden, wobei -2 für *auf keinen Fall* und +2 für *mit Sicherheit* stand. Der Fragebogen beinhaltete Fragen zu den folgenden Themen:

- Wie relevant waren die Fälle für Ihre ärztliche Tätigkeit?
- Wie fanden Sie die Art der Fallpräsentation am Rechner?
- Ihre Anmerkungen zur Gestaltung der Fortbildung?
- Ihre Anmerkungen zu den Fällen 1,2,3 und 4.

Die Bewertungen in den Fragebögen befanden sich meist in der Mitte zwischen null und eins (genaue Daten in [LEINER et al. 1999]). Grundsätzlich wurde das Computerprogramm als Ausbildungsmedium sehr gut angenommen, worauf auch die vielen Fallverlaufsprotokolle schließen lassen. Im Schnitt hat jeder Teilnehmer jeden Fall zweimal durchgearbeitet. Die aus den Fallprotokollen errechnete durchschnittliche Bearbeitungszeit lag bei etwa 25 Minuten pro Fall. Die Entscheidungshilfen wurden in allen Fällen nur selten benutzt. Zwischen den beiden Gruppen konnte kein deutlicher Unterschied festgestellt werden.

Bei den frei eingebbaren Kommentaren zu Fällen und Programm konnten die Ärzte ihre Meinung noch einmal äußern, wobei sie dies hauptsächlich zu medizinisch relevanten, also inhaltlichen, Aspekten taten. Generell bezeichneten die Ärzte die Bedienung des Systems im Vergleich mit Lehrbü-

chern als gewöhnungsbedürftig, fanden aber, daß die Lehrmethode zur aktiven Arbeit zwingt, was zunehmend interessanter wird.

Weitere Einsätze innerhalb der Qualitätszirkel sind geplant, wobei als nächster Schritt die Betreuung durch Tutoren geplant ist, die nicht an der Entwicklung der medizinischen Inhalte beteiligt waren. Als zusätzlicher Pluspunkt sei hier angemerkt, daß die Tutoren selbst eigene Fälle eingeben und in ihren Veranstaltungen benutzen können. Durch diese Option können sich die Tutoren mehr in den Kurs einbringen und sind nicht gezwungen, die vorgegebenen Fälle zu verwenden. Da sich die Diskussion über die medizinische Fallinterpretation als sehr wesentlich herausgestellt hat, könnte das Einbringen persönlicher Erfahrungen durch eigene Patientenfälle seitens der Tutoren für den Erfolg dieses neuen Lernmediums eine zentrale Rolle spielen.

4.4.1.3 GMDS–Kriterienkatalog

Die Fachgruppe Computer Based Training in der Medizin der GMDS entwickelte gemeinsam einen Kriterienkatalog für elektronische Publikationen in der Medizin.

Der Kriterienkatalog wurde auf den *RheumaTutor*, bzw. auf das *RheumaSeminar* angewandt von:

- Dr. Stefan Schulz, Abteilung Medizinische Informatik, Universität Freiburg (*RheumaSeminar*)[6]
- Dr. Florian Leiner, Abteilung Medizinische Informatik, Universität Göttingen (*RheumaTutor*)[7]
- Reinhard Kreutz, Institut für Medizinische Informatik und Biometrie, Universität Aachen (*RheumaTutor*)[8]

Dr. Schulz schreibt als Fazit:

> „Das auf der D3–Expertenshell basierende Rheumaseminar weist einen didaktisch hochinteressanten Ansatz auf, welcher an vier von einem ausgewiesenen Experten editierten Beispielfällen durchexerziert werden kann.

[6] Dr. Schulz ist von Haus aus Mediziner und arbeitet seit mehreren Jahren in der medizinischen CBT–Forschung. Er war an der Entwicklung des Kriterienkataloges intensiv beteiligt.

[7] Dr. Leiner ist Mediziner und war vor allem bei der Evaluation des *RheumaSeminars* innerhalb der Qualitätszirkel aktiv.

[8] Herr Kreutz ist Informatiker und promoviert im Bereich der medizinischen CBT–Systeme. Auch er war an der Entwicklung des GMDS–Kriterienkataloges beteiligt.

> Die derzeitige Realisierung hat noch prototypischen Charakter, da die Gestaltung der Benutzeroberfläche, die Bedienlogik und die für ein derartig komplexes System unabdingbaren Hilfen noch erhebliche Mängel aufweisen. Bei einem marktfähigen Produkt sollten zudem erheblich mehr Lernfälle zur Verfügung stehen.[9]
>
> Im Hinblick auf die Zielgruppe sollte die Einarbeitungszeit in das System in einem vernünftigen Verhältnis zum erwarteten Lernerfolg stehen. Daher muß auf die ergonomische Optimierung besondere Sorgfalt verwandt werden, was bei diesem System sicherlich noch einen erheblichen Aufwand erfordert. Insbesondere sollte dabei auf von Standardsoftware vertraute Bedienungselemente, Metaphern und Abläufe zurückgegriffen werden. In der gegenwärtigen Form ist das nicht der Fall. Zu einem Teil scheint dies auf die Entwicklungswerkzeuge zurückzuführen zu sein, zu einem weiteren Teil auf die Portierung von Macintosh nach Windows. Eine Lösung kann in einer kompletten Trennung zwischen Programmlogik und Bedienlogik liegen. Dies schafft dann die Möglichkeit, mit geeigneteren Programmierwerkzeugen (z. B. Delphi oder Visual Basic) eine adäquate Benutzeroberfläche zu entwickeln. Ebenso böte sich hierdurch eine Perspektive für die Entwicklung einer WWW–gestützten Client / Server–Version.“

Die beiden anderen Gutachter gaben kein eigenes Fazit ab, erwähnten aber die Tatsache, daß der Kriterienkatalog für das vorliegende Programm nicht vollständig passend sei. Die vollständigen Gutachten sind in Anhang 6.2 wiedergegeben.

Die Schwächen des D3TRAINERS sehen auch wir in der Bedienoberfläche, ein Problem, welches aber durch die große darzustellende Datenmenge und die gewünschte Offenheit gegenüber anderen Fachbereichen außerhalb der Medizin nicht trivial ist. Verschiedene Lösungsansätze sind aber vorhanden. Eine WWW–Client–Server–Version wurde innerhalb einer Diplomarbeit [FAULHABER 1997] bereits konzipiert und implementiert, d. h. eine Trennung von Benutzeroberfläche und den anderen Komponenten des D3TRAINERS ist bereits gegeben.

4.4.2 Neurologie

Der *Fallbasierte Neurologie–Trainer* wurde von Prof. Poeck entwickelt. Er bezieht sich genau auf das Wissen, das in seinem Neurologie–Lehrbuch

[9] Dr. Schulz lag die Version mit nur vier anstatt der 25 Lernfälle des kommerziellen Produktes vor.

abgedeckt wird. In dem Buch (verschiedene Auflagen: [POECK 1994] und [POECK und HACKE 1998]) befindet sich ein Gutschein, mit dem man das Lernprogramm kostenfrei beziehen kann. Im Gegensatz zu dem rheumatologischen System beinhaltet der Fallbasierte Neurologie–Trainer keine multimedialen Elemente zur Darstellung und Erkennung von Symptomen, sondern ist ausschließlich auf die Schlußfolgerungsfähigkeiten und Nachfragetechniken ausgerichtet [PUPPE et al. 1995].

Inzwischen (Stand 1. März 1999) wurde die Software (in der 2. Version) ca. 3000 mal in den verschiedenen Versionen (Windows 3.11, Windows95, Macintosh 68k, PowerMacintosh) verschickt. Dem Programm ist ein Fragebogen beigelegt (siehe Anhang **??**), von dem im Moment (ebenfalls Stand 1. März 1999) 102 Bögen zurückgeschickt wurden. Zusätzlich wurde das Programm von einzelnen Stellen (z. B. Springer Verlag, Universität Heidelberg, insgesamt sechs Fragebögen) getestet.

Im Sinne der in Abschnitt 4.1 vorgestellten Formen und Methoden der Evaluation handelt es sich hier um eine summative Evaluation im realem Einsatz. Da die Fragebögen durch die Systementwickler erstellt und ausgewertet wurden, liegt eine innere Evaluation vor, die das gesamte System betrachtet (Makroevaluation). Es handelt sich um eine nicht–vergleichende Evaluation. Als grundlegende Form sind hier natürlich die Fragebögen benutzt worden, wobei die Einbeziehung der gesondert ausgesuchten Stellen (Springer Verlag, Universität Heidelberg) als Expertenbefragung verstanden werden kann.

Das Ergebnis der vorliegenden 108 Fragebögen kann wie folgt zusammengefaßt werden: Der größte Teil der Benutzer, die den Fragebogen beantwortet haben waren Medizinstudenten (70), nur 20 Personen waren Ärzte im Praktikum oder in der Facharztausbildung, elf Benutzer waren bereits Fachärzte (für die Verteilung des Alters siehe Abb. 4.5). Bezüglich ihrer Computererfahrung (siehe auch Abb. 4.6) gaben die Benutzer Textsysteme, Tabellenkalkulation, Computerspiele und Graphikprogramme am häufigsten an. Nur 35 Benutzer hatten schon einmal mit medizinischen Trainingsprogrammen gearbeitet, Programmierkenntnisse konnten nur 25 Teilnehmer vorweisen.

Der überwiegende Teil der Benutzer gab an, das System ohne Probleme installieren zu können[10]. Fast allen (102) fiel die Einarbeitung in das System leicht, wobei etwa die Hälfte angab, die Bedienungsanleitung nicht oder kaum gelesen zu haben. 85 Benutzer waren mit der Bedienung des Systems zufrieden, wobei 46 davon Verbesserungsvorschläge machten. Am häufigsten von den beiden angebotenen Modi wurde die didaktische Variante gewählt, in der der Lernende die Frageklassen selbst anfordert und neben

[10]Es gab bei der ersten Version (März 1995) Installations–Probleme, da die Orginalinstallation auf Windows95 angepaßt werden mußte.

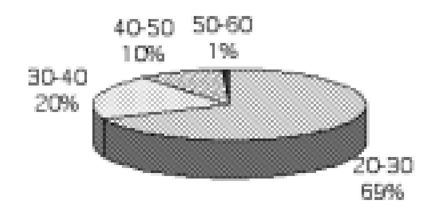


Abbildung 4.5: Altersstruktur bei den ausgewerteten 108 Fragebögen.

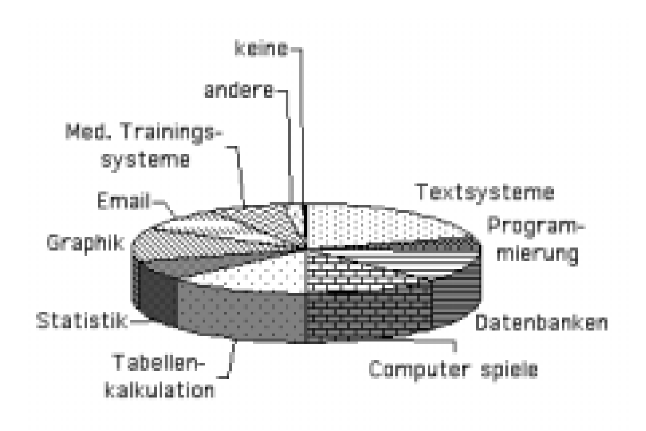


Abbildung 4.6: Frühere Computererfahrung bei den ausgewerteten 108 Fragebögen.

dem Stellen von Verdachtsdiagnosen auch Diagnose– und Testindikationsbegründungen angibt (54 Nennungen). An zweiter Stelle (47 Nennungen) wurde die Variante mit Symptomanforderung in Gruppen, Stellen von Verdachtsdiagnosen und Diagnosebegründungen genannt. Die Schwierigkeit der Fälle sowie das Feedback des Systems wurden durchwegs positiv beurteilt, so fanden 94 Benutzer die Trainingsfälle richtig und nur 10 Benutzer zu leicht. Das Feedback wurde von 41 Personen als gut bewertet, während 32 Benutzer *überwiegend gut, aber mit Ausnahmen* ankreuzten. Bei den Wünschen zur Weiterentwicklung zeigte sich folgende Gewichtung (siehe auch Abb. 4.7):

- Erweiterung auf Therapie (80)
- multimediale Patientenpräsentation (70)
- elektronische Kopplung mit Lehrbuch (61)
- detailliertere Kritik (51)
- elektronische Kopplung mit Lexikon (31)
- Betrachtung von mehr Diagnosen (26)
- Verbesserung der Bedienoberfläche (21)
- schwierigere Fälle (21)
- Andere Vorschläge (21)
- klinische Nutzung zur Kritik Ihrer eigenen Diagnostik bei neuen Fällen (20)
- klinische Nutzung für Konsultation bei neuen Fällen (18)

Insgesamt beurteilten die Benutzer das System grundlegend positiv. Die größten Schwierigkeiten bezüglich Laufeffizienz und Installation wurden durch neue Versionen beseitigt. Obwohl der Fallbasierte Neurologie–Trainer noch keine Bildsysteme anbietet und auch andere Optionen wie beispielsweise die Darstellung von Untersuchungskosten nicht benutzt, scheint das System für die Benutzer schon sehr attraktiv zu sein. Besondere Aufmerksamkeit bekommt der Fallbasierte Neurologie–Trainer aber natürlich durch die Verbindung mit dem weit verbreiteten Lehrbuch des Wissensbasisautors.

Bei dieser Anwendung liegt auch ein Aussage des Entwicklers über die aufgewendete Arbeitszeit vor. Der emeritierte Professor gibt an, über zwei Jahre durchschnittlich jeden Tag ungefähr 1 Stunde an der Wissensbasis und den 190 Fällen gearbeitet zu haben. Dabei wurde er von studentischen

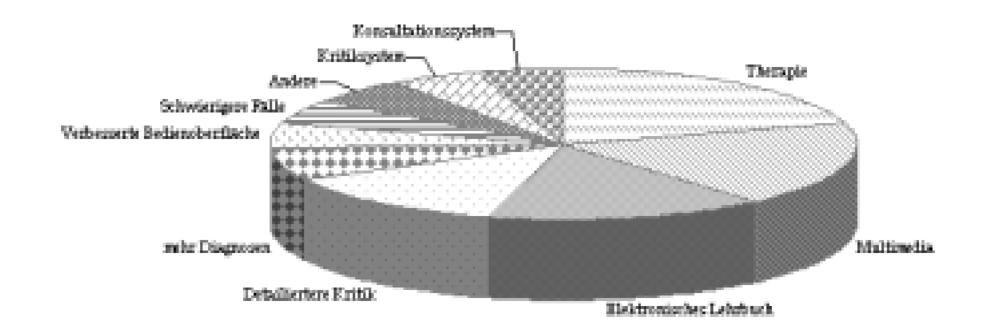


Abbildung 4.7: Wünsche der Anwender für die Weiterentwicklung des Neurologie–Trainers.

Hilfskräften in etwa gleichem Umfang unterstützt. Dies entspricht etwa 1000 Arbeitsstunden, also einem halben Personenjahr einer vollen Arbeitskraft. Die Wissensbasis umfaßt die Diagnostik der gesamten Neurologie entsprechend dem Lehrbuch, so daß die 190 repräsentativen Testfälle korrekt gelöst werden. Sie wurde jedoch nicht zur Lösung schwierigerer Fälle aus der Praxis optimiert.

Der Fallbasierte Neurologie–Trainer, wie er seit 1995 verschickt wird, beruht noch auf einer frühen Version des D3TRAINERS und wurde bisher nicht verändert. Die Erweiterung auf Therapien und die Anbindungen von Bildsystemen, wie durch die Benutzer gewünscht und in der aktuellen Version des D3TRAINERS ohne Probleme möglich, könnte eine wertvolle Weiterentwicklung des Programmes sein.

4.4.3 Hepatologie

In der medizinischen Klinik des DRK Krankenhauses Berlin–Köpenick wurde das von Prof. Buscher und seinen Mitarbeitern entwickelte Hepatologische Programm entwicklungsbegleitend evaluiert. Im Sinne der in Abschnitt 4.1 vorgestellten Aspekten liegt hier eine formative, innere Makroevaluation mit Laborcharakter vor, die sich hauptsächlich auf Fragebögen und Verhaltensbeobachtung stützt. Durch die verschiedenen Benutzergruppen bei dieser Studie kann ebenfalls von einer Expertenbefragung und einer vergleichenden Evaluation bezüglich der Anwendergruppen gesprochen werden.

Es wurden unterschiedliche Versionen von *HEPA–CADS*s in verschiedenen Entwicklungsstadien und mit unterschiedlicher Komplexität auf potentielle tutorielle Nutzbarkeit evaluiert. Dazu wurden die Testpersonen in drei

Gruppen eingeteilt:

- Gruppe I: Studenten im Praktischen Jahr (25 Teilnehmer)
- Gruppe II: Ärzte im Praktikum und in Weiterbildung zum Facharzt mit weniger als drei Jahren ärztliche Tätigkeit (15 Teilnehmer)
- Gruppe III: Fachärzte und Ärzte in Weiterbildung zum Facharzt mit mehr als drei Jahren ärztlicher Tätigkeit (11 Teilnehmer)

Die Testpersonen wurden einzeln in das Programm eingewiesen und hatten die Möglichkeit, innerhalb einer beliebig langen Sitzung exemplarische Fälle verschiedenster Schwierigkeitsgrade zu lösen. Die Befragung der drei Gruppen erfolgte mit Hilfe standardisierter Erhebungsbögen. Ein Teil der Informationen wurde vor dem ersten Kontakt mit dem Programm, ein Teil nach Kenntnis des Programms erhoben. Es wurden folgende Fragebögen von den Probanden ausgefüllt:

1. Fragebogen zu Lernverhalten und zu Erwartungen an ein computergestütztes Lernprogramm *vor* Kenntnis des Programms,
2. Fragebogen zur Akzeptanz von computergestützten Programmen *nach* Kenntnis des Programms.
3. Fragebogen zur Nutzbarkeit des kausalen Wissens des Programms *nach* Kenntnis des Programms,
4. Fragebogen zur Vorstellung über Art und Umfang der Erklärungskomponenten eines Lernprogramms *nach* Kenntnis des Programms,

Die drei Gruppen unterschieden sich in ihren Erwartungen an ein computergestütztes Lernprogramm in folgender Weise: Die Gruppen I und II wünschten Lernen anhand klinischer Fälle, eine fallbezogene Darstellung des zu vermittelnden Wissens, eine Tutorfunktion sowie eine differenzierte Darstellung der Diagnosen, Differentialdiagnosen und Symptome. Die Gruppe III wünscht dagegen ein Lernprogramm mit Funktion als Erinnerer im Sinne eines Drill– oder Testprogramms.

Die Mehrzahl der jüngeren Ärzte halten die Konsultation eines solchen Tutorprogramms für anregend, lehrreich und geeignet, Wissen mit Praxisbezug zu vertiefen (siehe auch Abb. 4.8). Ärzte mit Heimcomputererfahrung glaubten zudem überwiegend, daß selbständiges Denken geschult würde. Die älteren Ärzte glaubten dagegen häufig nicht, daß sie von einem Computerprogramm profitieren können. Alle drei Gruppen halten einen klinischen Lehrer für nicht durch ein Programm ersetzbar. Vor allem die jüngeren Ärzte würden Programme wie *HEPA–CADS* vorzugsweise in Ergänzung zum klinischen Unterricht einsetzen, insbesondere zur Nacharbeit eines in einem Untersuchungskurs vorgestellten Krankheitsfalles.

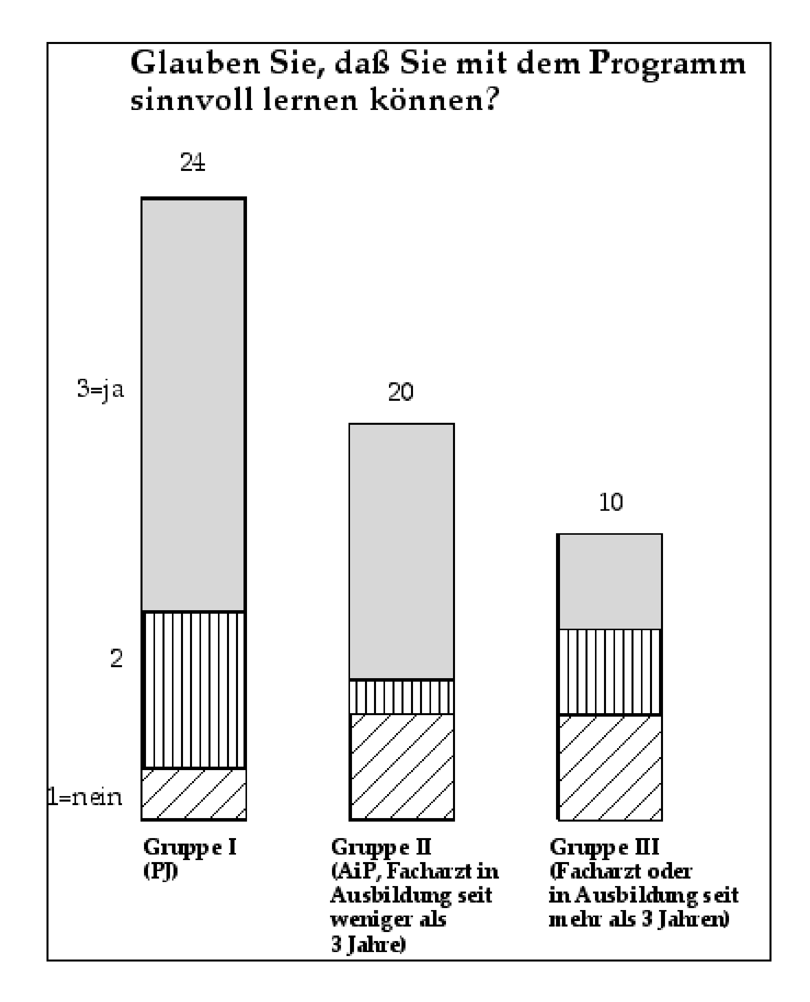


Abbildung 4.8: Darstellung der drei verschiedenen Gruppen und ihre Einschätzung, ob das präsentierte Programm HEPA-CADS sinnvoll erscheint.

Die genauen Daten der Befragung sind in Anhang **??** dargestellt.

4.4.4 Pflanzenerkennung

Im Rahmen einer Diplomarbeit der Biologie [Ernst 1996] wurde eine Wissensbasis zur *Erkennung von krautigen Blütenpflanzen* aufgebaut, die etwa 100 Pflanzen abdeckt [Reinhardt 1994]. Inhalt der Arbeit war nicht primär die tutorielle Anwendung, vielmehr sollte das Verhalten verschiedener Problemlösestrategien (heuristisch, überdeckend und fallvergleichend) miteinander verglichen werden.

Um die tutorielle Anwendung der Wissensbasis mit den multimedial aufbereiteten Fällen zu erproben, wurde im Sommersemester 1994 ein Kurs *Computergestütztes Einüben im Bestimmen von Blütenpflanzen* für Vordiplomstudenten der Biologie angeboten. Der Kurs war ursprünglich als freiwilliges Begleitangebot zum Bestimmungskurs des botanischen Instituts der Universität Würzburg geplant. Anhand eines Vergleichs zwischen den Klausurerfolgen der Studenten die den Computer verwendeten und denen, die ihn nicht verwendeten, sollte ein erster Eindruck über Brauchbarkeit dieses Lehransatzes gewonnen werden. Aus organisatorischen Gründen konnte der Kurs jedoch so nicht stattfinden, sondern mußte völlig unabhängig und räumlich getrennt durchgeführt werden. Es meldeten sich daher nur sechs Studenten, von denen nur drei den Kurs schließlich regelmäßig besuchten.

Eine sinnvolle Beurteilung des Lernerfolges ist daher nicht möglich, sieht man von der Aussage der Teilnehmer ab, daß das Arbeiten mit dem System für das Erkennen der wesentlichen Merkmale von Blütenpflanzen eine gute Hilfe gewesen sei.

Bei diesem Szenario kann zwar nicht von einer Evaluation gesprochen werden, sie zeigt aber dennoch, daß der D3Trainer auch erfolgreich außerhalb der Medizin angewendet werden kann. Das Pflanzenerkennungsprogramm soll in Zukunft in Zusammenarbeit mit verschiedenen Biologen bei einem Pflanzenerkennungskurs eingesetzt werden. Dazu werden die dort ansässigen Dozenten selbst Fälle und die entsprechenden Bildsysteme aufbauen und diese mit ihren Studenten besprechen.

4.5 Zusammenfassung Evaluation

Die Evaluation eines Systems durch Studien seiner Anwendung stellt nur einen Teil einer umfassenden Evaluation dar. Kriterienkataloge und Expertenbefragung sind sicherlich weitere Hilfsmittel, um die Qualität der eigenen Entwicklungen zu überprüfen und zu verbessern. Ein Überblick über diese Studien wird in Tabelle 4.3 gegeben. Im folgenden werden aber alle

Evaluation	**Anzahl Teilnehmer**	**Kommentar**
Rheumatologie Poliklinik München	104 Studenten	verschiedene Versionen mit Bildsystemen, Zeitraum von WS94/95 bis WS98/99
Qualitätszirkel	26+x niedergelassene und im Krankenhaus tätige Ärzte	vier bzw. 25 Fälle mit Bildsystemen, ab Sommer 1998
GMDS–Kriterienkatalog	-	25 Fälle mit Bildsystemen, Dez. 1998
Neurologie Fragebögen-auswertung	108 Fragebögen	frühe Version ohne Bildsysteme, Therapien und Kosten, aber hoher Fallanzahl (190), Fertigstellung März 1995
Hepatologie Krankenhaus Köpenick	51 Studenten und Ärzte	frühe Version ohne Bildsysteme, Therapien und Kosten, Evaluation im Sommer 1996
Pflanzenerkennung Biologie Würzburg	6 Studenten	ca. 10 Fälle mit Bildsystemen, Sommer 1996

Tabelle 4.3: Übersicht über die bisher durchgeführten Evaluationen.

Komponenten des D3TRAINERS nochmals diskutiert, um zu bestimmen, ob und wie sie ihre Aufgaben erfüllen. Als Basis für diese Diskussion dient die Tabelle 2.1, in der die Aufgaben und Anforderungen an ein intelligentes Trainingssystem beschrieben wurden.

Die Tatsache, daß sich die Falldarstellung aus Wissens–, Fallbasis und Bildsystemen generieren läßt, bringt den großen Vorteil mit sich, daß die Wartung und Erweiterung der Anwendungen problemlos durch den Fachexperten durchgeführt werden kann. Besonders durch die Option, auch die Bildsysteme zu generieren, können neue Fälle in kurzer Zeit bereitgestellt werden, was vor allem bei dem Einsatz an der Poliklinik München ein absolut notwendiges Kriterium war. Die zahlreichen Optionen bei der Interaktivität der Falldarstellung zeigten sich hauptsächlich bei der Betrachtung

mehrerer Anwendungen als wichtig. Eine einzelne Anwendung benutzte meist nur ein oder zwei der angebotenen Modi, über alle Anwendungen hinweg wurden aber alle Modi genutzt. Die detaillierte Beschreibung verbunden mit der Vielzahl der angezeigten Symptome erwies sich sowohl als vor– wie auch als nachteilig. Durch die große Masse an Daten und die somit entstehenden Probleme bezüglich der Übersichtlichkeit, die nur teilweise durch auf– und zuklappbare Hierarchien gelöst wurden, entstand oft eine erste Überforderung der Lerner, die sich erst mit häufigem Gebrauch des Systems auflöste. Der Vorteil dieses Detaillierungsgrads liegt in der Vermittlung der komplexen Sachverhalte, in dem rheumatologischen Bereich vor allem in der ausführlichen Durchführung der Anamnese. Durch diesen Ansatz wird die reale Komplexität des Falles nicht versteckt, sondern dem Lerner offensichtlich gemacht, so daß er auch später im praktischen Einsatz nicht durch die Masse an Falldaten überrascht wird. Hilfen bei dem Umgang mit den umfangreichen Falldaten bietet die Unterscheidung in abnorme und normale Symptome, wobei die normale Symptomatik auch vollständig ausgeblendet werden kann. Weitere Ansätze zur Unterstützung des Lerners müssen aber gefunden werden.

In den momentanen Anwendungen waren die verwendeten Medien Text und Bild ausreichend, allerdings kann man sich sinnvolle Einsätze weiterer Medien zur Symptomerkennung (z. B. Herztöne in der Kardiologie) oder als Hintergrundwissen (z. B. Animation eines schlagenden Herzens) gut vorstellen. Bei der Darstellung der Texte erscheint die schematisierte Präsentation der Falldaten als übersichtliche Variante, die aber auch noch durch graphische Hilfen unterstützt werden könnte. Bei dem Erklärungs- und Hintergrundwissen kann durch zusätzliche Formatierungsmöglichkeiten noch an Übersichtlichkeit gewonnen werden, da diese aber nicht im Mittelpunkt der Fallbearbeitung stehen, ist dies ein weniger interessanter Punkt, der nur zur optimalen Abrundung des Systems führen würde.

Die semioffene Frageform stellt eine optimale Umsetzung der Aktionsauswahl dar. Dem Lerner wird so die empfohlene Terminologie vermittelt, ohne ihm eine Hilfestellung durch eine zu kleine Vorauswahl zu geben. Ist die Bereitstellung mehrerer Synonyme aber unbedingt notwendig, muß dies bisher über die vorformulierte Erklärung zu einem Begriff erfolgen. Eine Freitexteingabe mit Nutzung der existierenden Synonymsammlungen wäre eine sinnvolle Erweiterung für diesen Bereich. Die angebotene MC–Auswahl für die Aktionsselektion ist in dem momentanen Status zu einfach für einen Studenten, da neben den richtigen Alternativen zufällig weitere Alternativen ausgesucht werden, die meist nichts mit dem momentanen Fall zu tun haben. Eine gesteuerte Sammlung von richtigen und ähnlichen oder naheliegenden Alternativen würde diese Option der Frageform deutlich aufwerten.

Die Navigation des Studenten durch den D3TRAINER über Buttons und

Menüs hat sich grundsätzlich als gute Lösung erwiesen. Für die einzelnen Anwendung wäre es allerdings durchaus eine reizvolle Option, wenn die Beschriftungen der Schaltflächen frei wählbar wären. So kann beispielsweise die Anfrage nach neuen Informationen in der Medizin *Untersuchungen anfordern* und in technischen Anwendung *Inspektion einzelner Teile* benannt werden. Innerhalb eines Themenbereiches, wie beispielsweise der Medizin, sollten aber sinnvolle Standards eingehalten werden, so daß die verschiedenen Anwendungen eines Bereichs einheitlich zu bedienen sind.

Ein in den Studien mehrmals erwähntes Problem ist eine anfängliche Gewöhnungsphase der Studenten bis diese mit dem System arbeiten können. Hier wäre neben den noch auszubauenden Kontextvisualisierungen auch Tooltips und eine Online–Hilfe zu nennen. Generell ist die Oberfläche des D3TRAINERS eine der Funktionalität angemessene Umsetzung, die auch bei den Studenten angenommen wird, durch einige Veränderungen aber noch optimiert werden kann.

Der Verzicht innerhalb des D3TRAINERS, den Wissenstand des Lerners zu modellieren und statt dessen nur Einzeleinstellungen und Stereotypen zur Modellierung der Präferenzen zu verwenden, hatte keine erkennbaren negativen Auswirkungen auf die Arbeit mit dem Studenten. Die dadurch fehlende automatische Fallauswahl wird durch komfortable Menüs für den Studenten ersetzt. So kann der Student seine Lernwelt selbst adaptieren, ohne durch das System in ein Schema gedrängt zu werden, welches nur in einem bestimmten Rahmen überhaupt auf ihn zutreffen kann. Hier gilt der Leitspruch ‚Besser kein Studentenmodell, als ein schlechtes'. Um aber ein gutes Studentenmodell zu bekommen, muß neben einem automatischen Monitoring auch eine immer wieder unterbrechende Eingabe erfolgen, die den normalen Fallverlauf stört. In den durchgeführten Studien zeigte sich, daß für die verschiedenen Anwendungen wirklich auch unterschiedliche Benutzerprofile erstellt wurden, die für eine Anpassung des Systems an die jeweiligen Bedürfnisse sorgte. Die Einführung von Stereotypen, die durch den Fachexperten definiert wurden, bietet eine leichtere Anpassung des Systems an den wenig geübten Benutzer, der evtl. anfangs noch keine Vorstellung von den Auswirkungen der einzelnen Einstellungen hat.

Der didaktische Grundgedanke hinter dem D3TRAINER, das problemorientierte Lernen zu unterstützen, konnte in den Anwendungen vollständig umgesetzt werden. Die Entscheidung für eine benutzergesteuerte Lernaktivität folgt daraus fast zwangsläufig. Trotzdem wurden einzelne Hilfen für den Studenten bereitgestellt, so daß er, auch ohne Hilfe eines menschlichen Tutors, den Fall eigenständig lösen kann. Die Beschränkung auf die kognitiven Aufgaben erfolgte zunächst aus technischen Gründen, da nur diese auf das basierende Expertensystem umgesetzt werden konnten. Eine Erweiterung auf manuelle Fähigkeiten durch die Kopplung mit Enablingsystemen oder organisatorischen Fähigkeiten durch die Integration in eine

virtuelle Lernwelt können aber mit der jetzigen Architektur umgesetzt werden. Kommunikative Fähigkeiten werden in dem D3TRAINER nur implizit (z. B. *Wie stelle ich eine Frage?*) oder explizit im Hintergrundwissen behandelt. Eine Erweiterung zu einer sozialen Simulation wäre jedoch durchaus denkbar.

Bei der Eingabe und Kritik der (kognitiven) Aufgaben des Studenten hat sich die momentane Version sehr stark gegenüber den Grundalgorithmen der ersten Version weiterentwickelt. Die Einführung von Wichtigkeiten und die Berücksichtigung ähnlicher Lösungen gaben den Ausschlag für eine benutzerorientierten Bewertung, wie sie im Moment vorliegt. Problematisch bleibt oft die Bewertung der Begründungen, die der Student für seine Aktionen eingibt. Bei der Begründung von Verdachtsdiagnosen wurde ein zweiter Problemlöser (überdeckend) herangezogen, da sich die vielstufigen, komplexen Regeln einer Konsultationswissensbasis als dem Lerner nur unzulänglich erklärbar erwiesen. Durch die wählbaren Problemlöser zur Erklärung ist dieses Problem nun aber auch gelöst. Schwierig bleibt weiterhin die Begründung zur Auswahl von Frageklassen, die nicht nach einem einfachem Algorithmus, sondern über eine Vielzahl miteinander verwobener Mechanismen begründet wird. Eine Lösung für dieses Problem ist noch nicht gefunden, so daß dem Fachexperten die Option angeboten wird, diese Begründungseingabe mit Bewertung auszublenden. Ansonsten wird das Feedback der Begründungen durch die Lerner gut angenommen und auch hier hat sich die semioffene Eingabeform bewährt.

Die kontextsensitiven Entscheidungshilfen, die dem Lerner im D3TRAINER bei einer Aktion die situationsabhängige richtige Lösung anzeigen, wurde in den Studien durch die Lerner sehr zurückhaltend benutzt. Diese Tatsache wird als positiver Hinweis darauf gesehen, daß sich die Lerner nicht gelangweilt durch den Fall führen lassen wollen, sondern das Interesse und den Ehrgeiz entwickeln, den Fall aktiv zu lösen und nur in Sackgassen nach der Lösung fragen. Die mehrstufige Lösungsnennung (z. B. erst die Bewertung der Benutzerdiagnosen, dann die Nennung der korrekten Diagnosen, anschließend erst die Hinweise auf diese Diagnosen) ermöglichen den Studenten, seine eigene Lösung zu entwickeln ohne die Lösungen aufgedrängt zu bekommen.

Der große Vorteil bei der Erstellung des Wissensmodells im D3TRAINER ist die Tatsache, daß bereits existierende Wissensbasen vollständig genutzt werden können, und die verschiedenen Optionen zur Mehrfachnutzung innerhalb des Expertensystembaukastens D3. Durch die verschiedenen Anbindungsmöglichkeiten für informelles und formales Wissen können alle Wissensarten, die in einem Trainingssystem nötig sind (strukturelles, strategisches, begriffliches und bildhaftes Wissen, Hintergrundwissen), auch in dem D3TRAINER verwendet werden. Die graphische Wissensakquisition für Wissens- und Fallbasis, sowie für die Bildsysteme, ermöglicht

es den Fachexperten ohne Programmierkenntnisse, ein vollständiges Wissensmodell eigenständig aufzubauen und zu pflegen. Bei der tutoriellen Nutzung von existierenden Wissensbasen innerhalb des D3TRAINERS traten einige kleinere Probleme auf, die zwar die grundlegende Funktionalität nicht beeinträchtigen, aber zu Schönheitsfehlern führten. Beispielsweise stellte sich heraus, daß für Konsultation und Training unterschiedliche Erklärungen benötigt wurden. Dieses Problem wurde durch die Einführung einer zweiten Erklärungsebene gelöst, welche optionale tutoriell sinnvolle Erläuterungen enthält.

Zur weiteren Unterstützung der Fachexperten bei der Wissenseingabe wurden Evaluationshilfen, wie automatisches Wissensbasistesten mit Fällen, eingeführt. Diese zusätzlichen Mechanismen ersparten den Experten viel Zeit und sind ein weiterer Aspekt bei der Qualitätssicherung.

Insgesamt stellt der D3TRAINER eine Implementierung eines Konzeptes dar, welches die grundlegenden Anforderungen an ein fallbasiertes Trainingssystem mit den Vorzügen der Generbarkeit verknüpft. Sowohl die Erstellung der Wissensmodelle, wie auch die Generierung der fallbasierten Trainingssysteme genügt den Ansprüchen dieser schwierigen Aufgabe. Die didaktischen Möglichkeiten wurden voll ausgenutzt und haben sich in den verschiedenen Anwendungen als sinnvoll erwiesen, wobei die Entscheidung für eine Variante in manchen Fällen durch den Wissensbasisautor und manchmal durch die Lernenden selbst gefällt wurden.

5

Ausblick

Die mit dem D3TRAINER entwickelten Trainingssysteme bieten eine intelligente Lernumgebung für das fallorientierte Lernen in diagnostischen Problembereichen. Vor allem in der problemorientierten Lehre können sie sehr gut eingesetzt werden, bieten aber auch für das Selbststudium außerhalb eines Curriculums genügend Unterstützung.

Der D3TRAINER vereint die Vorteile eines Autorensystems mit denen intelligenter Tutorsysteme. So kann ein Bereichsexperte ohne Programmierfähigkeiten selbständig ein Wissensmodell aufbauen und warten. Eine Tatsache, die sowohl praktische Vorteile hat, da der ‚Reibungsverlust' durch den Umweg über einen Wissensingenieur entfällt, wie auch motivationale Effekte, wie die größere Verbundenheit von Fachexperten und fertigem Produkt. Das Wissensmodell besteht in dem vorgestellten Konzept aus D3–Wissensbasis, einer dazugehörigen Fallbasis und eventuell Bildsystemen, die diese Fälle multimedial darstellen. Sowohl Wissensbasis wie auch die Fallsammlung können zusätzlich in einem Konsultationssystem (Second Opinion Programm) benutzt werden oder in anderen Bereichen (z. B. Falldokumentation, Informationssystem) mehrfach genutzt werden. Der Vorteil wissensbasierter Systeme wie dem D3TRAINER liegt in der hohen Flexibilität in Bezug auf die zulässigen Benutzeraktionen. In dem D3TRAINER kann der Lernende jederzeit alle Aktionen in beliebiger Reihenfolge durchführen, ohne daß das System seine Kritikfähigkeit verliert. Nur so kann eine aktive und eigenständige Fallösung durch den Lernenden geübt werden.

Die didaktische Komponente des D3TRAINERS besteht aus der Variation der grundlegenden Vorgehensweise in der Fallbearbeitung. Die angebotenen Möglichkeiten bezüglich Fallpräsentation, Kritikfähigkeit und Interaktivität können durch die Lernenden oder den Wissensbasisautor eingestellt werden und bieten somit eine flexible Umgebung für das problemorientierte Lernen. Hier setzt sich der D3TRAINER auch von anderen Systemen

ab, die nur eine didaktische Vorgehensweise anbieten oder nur geringe Variationen zulassen.

Auf die Integration eines Studentenmodells, wie es einige andere Systeme bereits benutzen, ist bei dem D3TRAINER bewußt verzichtet worden. Der Grund liegt in der Tatsache, daß aus den vorhandenen Daten, die bei einer Fallbearbeitung erhoben werden, nur schwer auf die internen Regeln geschlossen werden kann. Ein schlechtes, also nicht zutreffendes Studentenmodell, welches zur Adaption der Kritik oder zur Auswahl der Fälle benutzt werden könnte, bringt aber meist mehr Schaden als Nutzen.

Bei der Evaluation der Trainingssysteme wurde der größte Handlungsbedarf für Weiterentwicklungen bei der Benutzeroberfläche gesehen. Eine Online–Hilfe zur Unterstützung der Benutzer und andere Hilfesysteme (z. B. Tooltips) können schon elementare Verbesserungen bringen. Eine Möglichkeit, die in der jetzigen Version schon besteht, ist die Bereitstellung einer fachbezogenen Oberfläche für den jeweiligen Anwendungsbereich, die über die vorhandenen Schnittstellen kommunizieren. Eine Weiterentwicklung des WWW–Clients für die Nutzung der Systeme im Netz stellt ebenso eine wichtige Aufgabe für die Zukunft dar.

Der D3TRAINER in seiner momentanen Version bietet also eine sinnvolle Umgebung zur Generierung fallbasierter intelligenter Trainingssysteme im Bereich der diagnostischen Problemlösung. Die entstehenden Systeme können sowohl im Selbststudium, wie auch als Diskussionsbasis in der Gruppenarbeit benutzt werden, um aktive Fallösungen zu erlernen und einzuüben. Dabei ist es möglich, alle drei diagnostischen Aufgaben *Symptomerkennung*, *Testanforderung* und *Stellen von Verdachtsdiagnosen* zu unterstützen und zu bewerten. Die offene Architektur des D3TRAINERS ermöglicht außerdem den Aufruf elektronischer Lehrbücher.

Die durchgeführten Evaluationen zeigen, daß die generierten Systeme bereits erfolgreich eingesetzt wurden und ein sinnvolles Lernen ermöglicht wird. In der Zukunft sollen weitere Systeme erstellt und die vorhandenen Systeme verstärkt eingesetzt werden. In diesem Zusammenhang muß auch untersucht werden, inwieweit eine Kopplung zu anderen technischen Systemen möglich ist. Ein Beispiel für diese Zusammenarbeit ist die korrekte Handhabung eines Ultraschallgerätes, wenn diese Untersuchung in dem Trainingssystem angefordert wurde. Anstatt dem Studenten den entsprechenden Befund mitzuteilen oder das zugehörige Bild darzustellen, sollte er mit einem echten Ultraschallkopf an einer Puppe die richtige Untersuchung durchführen und das Ergebnis interpretieren können. Aber nicht nur eine Zusammenarbeit des D3TRAINERS mit solchen Enablingsystemen, sondern auch die Kopplung mit echten Laborgeräten bietet gute Ausbaumöglichkeiten für realitätsnahe Lernumgebungen. So muß der Lernende beispielsweise die Einstellungen an einem EKG–Gerät bestimmen und erhält das entstehende Bild zur Interpretation der Werte im D3TRAINER.

Die höchste Stufe der Realitätstreue wäre der Einsatz von Virtual–Reality–Methoden in einer Lernumgebung, die durch den D3TRAINER gesteuert wird und die auch in diesem Szenario alle Aktionen des Lernenden überprüfen und kritisieren kann.

Eine Erweiterung bezüglich der Handhabung von Therapien ist die Simulation der Therapieauswirkungen. Besonders in medizinischen Bereichen ist die Therapiestellung keine einmalige Aufgabe, sondern wie die Diagnose selbst, ein zeitabhängiger Ablauf aus Therapieauswahl, Beobachtung der Konsequenzen und Anpassung der Therapie. Für eine Therapiesimulation benötigt man die Repräsentation von Zeitverläufen, sowie ein Simulationsmodell der Therapieauswirkungen. Das Modellieren medizinischer Anwendungen ist wegen der hohen Komplexität des menschlichen Organismus allerdings sehr schwierig. Dabei hängt diese Komplexität aber von der Detaillierungsstufe des verwendeten Modells ab. Einige Erfahrungen auf dem Gebiet der simulationsbasierten Ausbildung in der Medizin werden in [ELIOT and WOOLF 1996] beschrieben. Unsere Erfahrung mit simulationsbasierten Anwendungen [FÄTH 1995] zeigen, daß die Erstellung des Modells eine schwere Aufgabe für den Bereichsexperten ist, die er oftmals nicht ohne Hilfe eines Wissensingenieurs bewältigen kann. Die Zusammenführung von diagnostischem Wissensmodell und Simulationsmodell für die Therapien stellt eine komplexe Aufgabe dar, die aber gelöst werden muß, soll die Therapiesimulation erfolgreich in die Fallbesprechung eingebaut werden.

Die durch den D3TRAINER aufgebauten Lernumgebungen sehen als pädagogischen Schwerpunkt das Anwenden und Einüben von Wissen, welches zuvor in einer anderen Lernumgebung (Vorlesung, Lehrbuch) erworben wurde. Es wird zwar die Möglichkeit angeboten, aus dem D3TRAINER heraus einzelne Seiten oder ganze elektronische Lehrbücher anzubinden, diese Verknüpfung ist aber nur nach außen gerichtet, d. h. im D3TRAINER ist nicht bekannt, welches Wissen sich der Lernende angeschaut hat. Durch die bidirektionale Verknüpfung der aktiven Lernumgebung des D3TRAINERS mit einem elektronischen Lehrbuch können aber viele didaktische Vorteile entstehen. Wie in [SETTELE 1998] konzipiert, kann ein Studentenmodell gemeinsam benutzt werden, in dem die gesamte Historie des Lernenden gespeichert wird. Jetzt kann das Trainingssystem auf das Vorwissen des Lernenden eingehen, also beispielsweise nur solche Fälle aussuchen, die dieser bereits lösen kann oder dessen Grundlagen er als letztes bearbeitet hat. Andererseits kann das Lehrbuch auf Schwachpunkte in der Fallösung des Studenten eingehen und ihm entsprechende Kapitel des Lehrbuchs empfehlen oder mittels einer adaptiven Oberfläche zumindestens einen Überblick über sein Wissen geben.

Eine Sichtweise, die in der Lehre immer stärker vertreten wird, ist die Integration der Diagnostik in die Gesamtumgebung, in der der Lernende spä-

ter sein Wissen anwenden soll. In dem medizinischen Bereich kann beispielsweise nicht davon ausgegangen werden, daß die einzige Aufgabe des Arztes sein wird, den Patienten korrekt zu diagnostizieren. In der realen Anwendung muß der Arzt beispielsweise auch nicht nur einen sondern mehrere Patienten behandeln und dies somit in seiner Zeiteinteilung berücksichtigen und aufeinander abstimmen. Ebenso muß er in seiner Zeiteinteilung auch beachten, daß er sich mit anderem medizinischen Personal abstimmen muß (andere Ärzte, Pflegepersonal und Nutzung von Untersuchungsgeräten). Nicht zuletzt sollte der Arzt, wenn möglich, auch auf die Wünsche des Patienten eingehen. Um später im Berufsleben erfolgreich sein Wissen einsetzen zu können, muß er auch auf diese Aufgaben vorbereitet werden. Viele dieser Aufgaben interagieren aber auch miteinander. So hängt die Auswahl der Station, auf die der Patient gelegt wird, hauptsächlich – aber nicht ausschließlich – mit der Diagnose des Patienten zusammen. Weitere Faktoren sind die ausgewählte Therapie, die Verfügbarkeit von Betten in dieser Station und andere nicht–medizinische Aspekte. Szenarien, in denen versucht wird, dieser Komplexität gerecht zu werden, nennt man virtuelle Umgebungen [SEITZ et al. 1999]. Auch hier ist es denkbar und wünschenswert, den D3TRAINER für die diagnostische Fallbearbeitung in ein solches Szenario zu integrieren. Angedacht ist beispielsweise die Kopplung mit dem Multiagentensystem SeSam [KLÜGL and PUPPE 1998], welche ermöglicht, den D3TRAINER als eigenständigen Agenten in die Simulationsumgebung aufzunehmen.

Die letzte Möglichkeit für eine Weiterentwicklung des D3TRAINER–Konzeptes soll hier die Anwendung als Kritikumgebung während des realen Einsatzes sein. Die Grundidee ist den Diagnostiker bei der Arbeit zu beobachten (Monitoring) und ihn nur dann zu unterbrechen, wenn ein berechtigter Einwand bei seiner Vorgehensweise vorliegt. Das Kritiksystem muß sehr sensibel reagieren, denn bei wichtigen Fehlentscheidungen muß es unbedingt eingreifen, darf den Benutzer aber nicht fälschlich kritisieren, da es sonst unglaubwürdig erscheint. Das gesamte Szenario kann auch nur dann funktionieren, wenn die Fallbearbeitung elektronisch überwachbar ist, d. h. wenn die Falldaten (Symptome) und die Diagnose– und Therapieauswahl des Diagnostikers dem System verfügbar sind. Erreichbar wäre dies durch die Integration in das Dokumentationssystem des Anwendungsbereiches, also beispielsweise der elektronischen Krankenakte oder einer elektronischen Fehlerdokumentation des Monteurs in technischen Domänen. Für den sinnvollen Einsatz solcher Systeme ist allerdings Zusatzwissen nötig [PUPPE 1998], welches vor allem die Vollständigkeit und Sicherheit der Daten und des Wissens sowie die Konsequenzen von Diagnosen und Therapien betreffen. So ist beispielsweise die Angabe eines Schmerzgrades (geschätzt von 1 bis 10) durch den Patienten weniger verläßlich als die gemessene Körpertemperatur. Andererseits soll das System

nicht eingreifen, wenn eine ungefährliche Diagnose, beispielsweise Grippe, diagnostiziert wurde, aber durch das System bezweifelt wird. Es ist aber unbedingt nötig einzugreifen, wenn eine Operation angeordnet wird und gewisse Falldaten (z. B. Blutdruck) stark dagegen sprechen. Das Kritiksystem kann natürlich nicht vollständig aus dem D3TRAINER abgeleitet werden, kann aber viele Funktionalitäten und Kritikmechanismen übernehmen.

6

Anhang: Originaldaten zur RheumaTutor-Evaluation

6.1 Daten der Studie an der Poliklinik

Im folgenden sind die Mittelwerte und Standardabweichung der zusammengefaßten Studien an der Poliklinik dargestellt. Die Anzahl der Teilnehmer beträgt somit insgesamt 82 Studenten.

Die Skala der Werte geht von 0 (schlecht) bis 10 (sehr gut). In der Tabelle werden Mittelwert, Standardabweichung und die Anzahl der Fälle angegeben, in denen diese Fragen beantwortet wurden.

Frage	Mittel–wert	Stand. abw.	Anzahl Fälle
Wert der Rheumatologie im Studium	6,61	1,68	82
Wert der Rheumatologie im späteren Beruf	6,22	2,14	82
Technische Probleme	3,35	2,68	81
Selbsteinschätzung des Wissens vorher	3,10	1,42	82
Selbsteinschätzung des Wissens nachher	5,69	1,51	80
Tutorprogramme sind sinnvoll	7,61	1,94	82
Dieses Programm ist optimal für CBT	6,07	1,87	82
Einsatz innerhalb des Kurses	6,44	2,34	81
Einsatz an PC in der Universität	5,01	2,66	82
Einsatz zu Hause	7,30	2,58	81
Einsatz innerhalb einer Gruppe	3,52	2,49	80

Frage	Mittel–wert	Stand. abw.	Anzahl Fälle
Einsatz zum Selbststudium	7,42	2,32	82
Einsatz zusätzlich zum Kurs	6,99	2,56	82
Einsatz als Ersatz des Kurses	1,55	1,21	82
Einsatz als Vorbereitung für eine Prüfung	5,58	2,70	82
Einsatz für Ausbildung in einem Spezialgebiet	6,61	2,67	82
Motivierung für unbekannte Fachbereiche	6,05	2,74	60
Lernen mit dem Programm ist sinnvoll	6,55	2,32	82
Arbeiten mit dem Programm ist nicht langweilig	6,88	2,06	82
Einsetzbarkeit für Fallanalyse	6,81	1,99	80
Anwendbar, um Begriffe zu klären	5,59	2,30	81
Qualifikation, um Anamnese zu erlernen	6,00	2,43	81
diagnostische Fertigkeit erlernen	6,73	2,29	81
Lerntempo	4,53	2,36	81
Unterstützung der Diagnostik	6,30	2,28	81
Lernmotivation	6,31	2,43	81
Diagnostischer Ablauf	5,63	2,56	81
Wissenstransfer	6,76	2,13	81
Verbindung zum praktischen Einsatz	5,09	2,68	81
Ersatz für menschlichen Dozenten	1,79	1,38	81
Unterstützung des Dozenten	6,86	2,25	81
Training von Faktenwissen	4,85	2,31	81
selbständiges Denken	6,36	2,81	81
Verständlichkeit der Texte	7,17	1,63	77
Strukturierung der Texte	6,71	1,92	77
Lesbarkeit der Texte	6,32	2,24	77
Graphische Präsentation	5,49	2,23	77
leicht zu erlernen	7,26	2,10	77
Robustheit	5,34	2,82	77
Klarheit	5,47	2,34	77 .
weiter mit dem Programm lernen	6,43	1,96	77
Verständlichkeit	6,78	2,04	77
Erklärung der eigenen Hypothese	5,66	2,32	77
Spaß, mit dem Programm zu arbeiten	6,51	2,06	77
Konzentration auf die Arbeit	5,88	2,28	77

Frage	Mittel–wert	Stand. abw.	Anzahl Fälle
Brauchbarkeit für das Studium	6,84	2,37	77

6.2 Begutachtung nach GMDS–Kriterienkatalog

Im folgenden wird der *RheumaTutor*, bzw. das *RheumaSeminar* anhand des GMDS–Kriterienkataloges begutachtet. Der Katalog wurde als Bewertungsgrundlage für elektronische Publikationen in der Medizin erstellt. Obwohl er nicht vollständig für die Bewertung solcher dynamischen, generierten Systeme wie dem D3TRAINER geeignet ist, so kann doch aus den Aussagen der Gutachter wertvolle Hinweise gezogen werden. Das *RheumaSeminar* unterscheidet sich vom *RheumaTutor* hauptsächlich durch die Anzahl der Fälle, so besitzt das Seminar nur vier Fälle im Gegensatz zu den 25 Fällen des Tutors, der auch alle verlagsüblichen Auflagen (z. B. ISBN–Nummer) erfüllt. Das *RheumaSeminar* wurde speziell für die Studien innerhalb der Qualitätszirkel erstellt, deshalb auch die automatische Protokollierung des Fallverlaufs auf Diskette.

Der Kriterienkatalog wurde auf den *RheumaTutor* bzw. auf das *RheumaSeminar* angewandt von:

- Dr. Stefan Schulz, Abteilung Medizinische Informatik, Universität Freiburg (*RheumaSeminar*)[1]
- Dr. Florian Leiner, Abteilung Medizinische Informatik, Universität Göttingen (*RheumaTutor*)[2]
- Reinhard Kreutz, Institut für Medizinische Informatik und Biometrie, Universität Aachen (*RheumaTutor*)[3]

Im folgenden wird der Kriterienkatalog und die gegebenen Antworten wiedergegeben. Nicht relevante Kriterien, zu denen keine Aussagen gemacht wurden, sind nicht aufgeführt. Die Meinungen der verschiedenen Gutachter wurden zusammengefaßt und in *schräger Schrift* dargestellt. Soweit sich die Meinungen der Gutachter relevant unterscheiden, werden die Meinungen mit den Namen zusammen dargestellt. Meine eigenen Anmerkungen zu den Punkten sind in `gesonderter Schrift` dargestellt.

[1] Dr. Schulz ist von Haus aus Mediziner und arbeitet seit mehreren Jahren in der medizinischen CBT–Forschung. Er war an der Entwicklung des Kriterienkataloges intensiv beteiligt.

[2] Dr. Leiner ist Mediziner und war vor allem bei der Evaluation des *RheumaSeminars* innerhalb der Qualitätszirkel aktiv.

[3] Herr Kreutz ist Informatiker und promoviert im Bereich der medizinischen CBT–Systeme, auch er war an der Entwicklung des GMDS–Kriterienkataloges beteiligt.

Qualitätskriterien für Elektronische Publikationen in der Medizin der GMDS

2.1. Inhalte

2.1.1. Autorenschaft

- Die medizinischen Inhalte werden korrekt und umfassend vermittelt.
  [Schulz:] Ja, soweit ich das als Nicht–Rheumatologe beurteilen kann. Manches könnte noch etwas weiter spezifiziert sein. Z. B. fehlen unter ‚Gichttherapie' Medikamente und Dosierung.

- Persönliche Meinungen der Autoren sind als solche kenntlich gemacht.
  Ja.

- Fachautoren, Herausgeber, Zeitangabe, Versionsnummer sind benannt und dokumentiert.
  Autor und Zeit, Herausgeber ja, Versionsnummer nicht.

- Zitate und herangezogene externe Quellen sind als solche kenntlich gemacht.
  Ja.

- Bezüge zu in der Medizin gebräuchlichen Ordnungssystemen sind hergestellt.
  [Schulz:] Nein.

- Bei Verlagsprodukten geht der Herausgabe ein internes (Lektorat) oder externes (Gutachter) Review–Verfahren voraus. Systeme zur Entscheidungsunterstützung sind in der Praxis evaluiert.
  `Der RheumaTutor wurde einem Lektorat durch den Verlag Ullstein Medical unterzogen. Das Wissensmodell zur Entscheidungsfindung, welches dem System zugrunde liegt, wurde in mehreren Studien anhand von realen Fällen aus der Praxis überprüft.`

2.1.2. Formale Anforderungen

- Verlagsprodukte sind in Katalogen des Buchhandels verzeichnet und mit einer ISBN–Nummer versehen.
  RheumaTutor ISBN 3-86126-934-1

- Angaben zum Copyright sind vorhanden.
  Ja.

- Die Lizenzbestimmungen enthalten zum Mehrbenutzerbetrieb und zum Verleih klare Aussagen.

```
Mehrbenutzerbetrieb nicht möglich. Zum Verleih
werden keine Aussagen gemacht.
```

2.1.3. Zielgruppenbezug

- Art und Umfang des notwendigen DV–Wissens sind deutlich ausgewiesen.
  Nein
  ```
  Da hier kein DV-Wissen vorausgesetzt wird, wird
  dies nicht extra erwähnt.
  ```

- Umfang und Tiefe des enthaltenen Wissens sind deutlich ausgewiesen.
  [Schulz:] Nein
  ```
  Siehe Bedienanleitung S.10-12, lag aber Dr. Schulz
  leider nicht in der endgültigen Version vor.
  ```

- Bezüge zu externen Lernzielkatalogen (z. B. GK1 – GK3 im Medizinstudium) sind hergestellt.
  [Schulz:] Nein.
  ```
  Aus dem Benutzerhandbuch(S. 12): „RheumaTutor
  soll weder Fach- noch Lehrbücher ersetzen.
  Es soll vielmehr praktisch tätigen Ärztinnen
  und Ärzten die Möglichkeit geben, sich anhand
  ausgewählter, geeignet dargestellter und
  interaktiv bearbeitbarer Beispielfälle aus der
  Rheumatologie fortzubilden.“
  ```

2.2. Technik

2.2.1. Plattform

- Die Anwendung ist für die bei der Zielgruppe zu erwartende Systemumgebung entwickelt. Im Idealfall werden mehrere gängige Systeme bezüglich der gesamten Funktionalität unterstützt.
  Es werden Bedenken wegen der Systemvoraussetzung von 32MB geäußert.

- Die Systemvoraussetzungen und –einschränkungen sind deutlich spezifiziert.
  Ja.

- Die Anwendung ist nur dann an bestimmte Bildschirmeinstellungen (feste Auflösung oder Farbtiefe) gebunden, wenn der Inhalt dies erzwingt.
  [Schulz:] Es gibt keinen inhaltlichen Anlaß für 32000 Farben. Die Forderung nach dieser Farbtiefe läßt sich mit derzeit gängigen Terminal–Lösungen nicht ohne weiteres vereinbaren.

```
Das Erkennen von medizinischen Symptomen in
Bilder kann nur durch Bilder mit hoher Qualität
unterstützt werden. In der Rheumatologie
sind oft Hautveränderungen wichtig für die
Diagnosestellung, die aber nur in hochwertigen
Bildern erkannt werden können. Aus diesen Gründen
ist eine hohe Farbtiefe unbedingt notwendig.
```

- Mehrbenutzerbetrieb auf Multi–User–Betriebssystemen ist berücksichtigt.
  [Kreutz:] Nicht beabsichtigt.
  [Schulz:] Es scheint zu funktionieren, ist aber nirgends dokumentiert. Parallele Schreibzugriffe auf dieselbe Lisp–Image-Datei werden sicherlich zu Problemen führen.
  ```
  Der D3Trainer ist in dieser Ausführung nicht für
  Multi-User-Betriebssysteme ausgerüstet. Für diese
  Anwendung wird auf die Client-Server-Anwendung
  verwiesen.
  ```

2.2.3. Installation / Erstaufruf

- Die Anwendung ist direkt vom Datenträger ohne Setup–Routine lauffähig.
  Ja.

- Die Anwendung ist auch ohne Eingriff in die Systembereiche des Betriebssystems lauffähig.
  Ja.

- Die Anwendung erfordert nach der Installation keinen System–Neustart oder manuelle Konfigurationsänderungen.
  Ja.

2.2.4. Performanz / Flexibilität / Laufzeitverhalten / Schnittstellen

- Die Anwendung ist stabil, robust gegen Bedienungsfehler, zuverlässig und performant.
  Keine Instabilitäten aufgetreten.

- Wo lange Antwortzeiten nicht zu vermeiden sind, wird der Nutzer durch Warnhinweise unterrichtet.
  [Schulz:] Leider nur teilweise, beispielsweise beim Starten des Systems.

2.3. Informationskodierung und Präsentationsmodalitäten

2.3.1. Text

2.3.1.1. Metainformationen

- Das Gesamttextvolumen der Publikation ist klar erkennbar.
  [Kreutz:] Ja.
  [Schulz:] Nein.
  Da die Texte hier nur erläuternden Charakter haben und nicht die eigentliche Funktionalität des Programmes betreffen, werden diese nicht gesondert erwähnt. Das Gesamtvolumen der Publikation, die 25 zu lösenden Fälle, sind deutlich genannt.

- Metainformationen, wie Autoren, Version, evtl. Zusammenfassung, lassen sich von überall einsehen, ohne daß dazu die aktuelle Seite verlassen werden muß.
  [Kreutz:] Nicht nötig.
  [Schulz:] Nein
  Alle erklärende Texte haben im Titel die Autoren oder sonstige Quellen genannt. Weitere Metainformationen werden nicht benötigt.

2.3.1.2. Formale Aspekte

- Inhalte sind prägnant und knapp formuliert.
  [Kreutz:] Ja.
  [Schulz:] Was die medizinische Inhalte betrifft, größtenteils. Semantische Unschärfen: Um Laboruntersuchungen anzufordern, muß ich teilweise Laborergebnisse anfordern. Beim Begrüßungs–Fenster: „Ihnen stehen folgende Informationen zur Verfügung: Inspektion beider Hände". Von der Bedeutung her unklar. Inspektion ... ist keine Information, sondern eine Untersuchung, mittels derer ich Informationen gewinne. Diese Information steht mir zu diesem Moment aber nicht zur Verfügung. Die sprachliche Form von Dialogen ist manchmal etwas ungenau (s.u.).

- Die Texte sind stilistisch, orthografisch und grammatikalisch korrekt. Die Interpunktion ist regelgerecht.
  [Kreutz:] Ja.
  [Schulz:] Geringe orthographische Ungenauigkeiten, z. B. System Fenster statt Systemfenster oder System–Fenster, Magen Darm Probleme statt Magen–Darm–Probleme etc. (sogenannte complex noun phrases sind im Englischen die Regel, im Deutschen eigentlich nicht erlaubt) Patientenantworten: nach Gleichheitszeichen mal groß, mal klein geschrieben, mal mit Ausrufezeichen, mal ohne. ‚Welche Gelenkveränderungen wurden im einzelnen durch den Arzt objektiv festgestellt' (sind mit ‚Arzt' die Benutzer gemeint ?)

2.3.1.3. Layout

- Es herrscht Konsistenz bezüglich Layout, Schriftart und Formatierung.
  [Kreutz:] Ja.
  [Schulz:] Die Schriften sind weithin einheitlich, in den Hilfetexten allerdings zu klein. Außerdem wird eine Formatierung vermißt (zumindest hervorgehobene Überschriften und Gliederung in Absätze). Diese sehr sparsame Textgestaltung fällt insbesondere auf, als beim Impressum und beim Begrüßungsfenster wesentlich ansprechendere Gestaltungselemente verwendet werden . . .
  `Die erklärenden Texte sind reine Zusatztexte, die nicht im Mittelpunkt des Systems stehen, in der nächsten Version werden diese aber als HTML-Texte mit entsprechender Formatierung angebunden werden.`

- Die Bildschirmfenster erscheinen niemals „vollgeschrieben". Das Scrollen von Text wird weitgehend vermieden.
  [Kreutz:] Ja, bis auf das ‚Message Fenster', welches sehr breit und ohne Zeilenumbruch ist.
  [Schulz:] Sehr häufig müssen Textfenster gescrollt werden, weil der Inhalt in einem unnötig großen Schriftgrad dargestellt ist (z. B. Fenster mit den Anfangsinformationen zum Fall). Bei ‚Erklärende Texte' muß sogar ohne Notwendigkeit waagerecht gescrollt werden muß.
  `In dem D3Trainer wurde immer versucht, die Fenster so groß zu machen, daß normalerweise nicht unnötig gescrollt werden muß. Auch hier muß erwähnt werden, daß die erklärenden Texte bei der grundlegenden Konzeption nicht berücksichtigt wurden, sondern als späterer Zusatz dazukamen und deshalb nicht optimal gestaltet werden konnten. Die wesentlichen Teile des D3Trainers jedoch, also die reine Fallbearbeitung, wurde auf diese Aspekte hin ausgerichtet.`

- Die Regel „Ein Thema – ein Textfenster" ist weitgehend realisiert.
  [Kreutz:] Ja.
  [Schulz:] Der Sinn der Trennung in die Fenster ‚Symptome des Falls' und ‚Wählen Sie bitte eine oder mehrere Informationen aus' ist nicht klar.
  `Das Fenster ‚Symptome des Falls' stellt die momentane Fallsituation dar, während das Fenster ‚Wählen Sie bitte eine oder mehrere Informationen aus' zur Auswahl neuer Untersuchungen dient, die dann in zur Veränderung der Fallsituation führen. In diesem Sinne ist pro Fenster eine Aufgabe der`

`Fallbearbeitung dargestellt und die Regel (ein Thema - ein Textfenster) konsequent umgesetzt.`

2.3.1.4. Abkürzungen, Fachtermini

- Der Gebrauch von Abkürzungen ist auf die umgangssprachlich bzw. im Fachgebiet üblichen zu beschränken.
  [Kreutz:] Ja.

- Selten benutzte Abkürzungen und Fachtermini werden mit Hilfe eines Lexikons oder Glossars erläutert. Ein Glossar ist benutzerspezifisch erweiterbar.
  [Kreutz:] Nein.
  [Schulz:] Ein Glossar wird vor allem bei manchen Laborparametern vermißt.
  `Zu jedem Laborparameter gibt es eine Erklärung, die durch Klick darauf aufgerufen wird.`

2.3.2. Grafik, Animation, Filmsequenzen, Fotografien, Tondokumente
2.3.2.1. Grafiken und Fotografien

- Die Farb– und Detailtreue fotografischer Darstellungen wird bestimmten Mindestanforderungen gerecht, die durch die bei der Zielgruppe vorhandenen technischen Ausstattungen erfüllt sind. Trotz hierdurch bedingter Kompromisse sind die Abbildungen aussagekräftig.
  [Kreutz:] Ja.

- Eine ‚Zoom '– Funktion bringt echte Zusatzinformation und beschränkt sich nicht auf eine Vergröberung der Bildauflösung.
  `In den Symptomerkennungen können mehrere Ansichten eines Bildes, also auch echte Vergrößerungen angesehen werden.`

- Bilder sind mit einer eindeutigen Kennung versehen, die kontextunabhängig ist.
  [Kreutz:] Ja.
  [Schulz:] Nein.
  `Die Bilder sind nicht als eigentlicher Teil der Wissensvermittlung verwendet, sondern sind in die Fallbearbeitung eingebunden, d.h. da sie gar nicht außerhalb der Fallumgebung aufgerufen werden können, müssen sie auch für den Benutzer keine kontextunabhängige Kennung aufweisen.`

2.4. Ergonomie und Design

2.4.1. Grundanforderungen

- Für die Bedienung sind keine DV–technischen Spezialkenntnisse erforderlich.
  Ja.

- Die Anwendung kann von jeder Stelle aus jederzeit beendet werden.
  [Kreutz:] Ja.
  [Leiner:] Ja.
  [Schulz:] Nein.

- Die Benutzungsoberfläche ist an gängigen GUI–Standards orientiert, die den Benutzern vertraut sind.
  [Kreutz:] Ja.
  [Leiner:] Ja.
  [Schulz:] Nein.

- Die Grundfunktionalität erschließt sich auch ohne vorherige Schulung oder Konsultation von Hilfetexten.
  Nicht ohne weiteres.

- Mausfunktionen sind auch mit Tastaturkombinationen auslösbar.
  Nicht durchgängig, manchmal gar nicht.
  `Der D3Trainer wurde daraufhin konzipiert, daß er nur mit der Maus und wenn möglich ohne Tastatur bedienbar ist.`

- Selten benötigte Funktionen sind über die Standardmenüleiste oder über Zusatzfenster abrufbar.
  Ja.

2.4.2. Steuerelemente

- Icons und Schaltflächen verwenden plausible Metaphern.
  Keine Icons, sondern nur Beschriftungen, die manchmal nicht vollständig lesbar sind (z. B. Aufklap . . .).
  `Da der D3Trainer domänenunabhängig einsetzbar bleiben soll, ist es sehr schwer, solche aussagekräftigen Icons zu finden. Das Abschneiden der Texte wird in der neuen Version vermieden werden.`

- Die Funktion eines Steuerelements ist verständlich.
  [Kreutz:] Ja.
  [Schulz:] Oft nicht, sog. Tooltips wären hilfreich.

- Merkmale der von der Zielgruppe im Alltag eingesetzten Standardsoftware (Officepakete, Web–Browser, Mailprogramme, Betriebssystemfunktionen etc.), sind, soweit übertragbar, in die Gestaltung der Benutzerschnittstelle mit einbezogen. Dies betrifft Funktionen wie:
  - Speichern benutzerspezifischer Einstellungen
    Ja.
  - Dateioperationen
    Standarddialog, funktioniert aber nicht wie erwartet.
    `Der im D3Trainer verwendete Dialog ist der Standarddialog ohne Veränderung.`
  - Maustastenbelegung
    Keine Funktion der rechten Maustaste.
- Anzahl und Vielfalt der Steuerelemente sind auf das notwendige Maß beschränkt.
  Ja.
- Steuerelemente finden sich immer an der gleichen Stelle und haben im gesamten Programm dasselbe Erscheinungsbild.
  Ja.
- Inaktivierte Steuerelemente bleiben weiterhin sichtbar, sind aber als inaktiviert eindeutig zu erkennen.
  [Kreutz:] Ja.
  [Schulz:] Manche Menus sind aktiviert, haben aber keine Funktion in dieser Situation.

2.4.3. Raumaufteilung

- Die Raumeinteilung des Bildschirms ist übersichtlich.
  [Kreutz:] Ja.
  [Leiner:] Könnte noch optimiert werden.
  `Ein Aufteilung durch Karteikästen ist geplant.`
- In standardisierten Bereichen ist prinzipiell immer die gleiche Art von Information zu finden.
  [Kreutz:] Ja.
  [Leiner:] Ja. Die Fenster unterscheiden sich damit aber auch zu wenig, so daß der Lerner Orientierungsprobleme bekommen kann.
  [Schulz:] Die Fenster haben aber nicht immer charakteristische Namen, . . .
  `Die Lösung dieses Problems soll durch charakteristische Hintergrundbilder gelöst werden, so daß der Lerner zwar dieselben Standardaktionen in den Fenstern durchführen kann, aber durch das`

`Bild immer weiß, in welchem Bereich er sich gerade befindet.`

- Eine Vielzahl von offenen Fenstern wird vermieden.
  [Kreutz:] Ja.
  [Leiner:] Im wesentlichem ja, könnte aber noch optimiert werden.
  [Schulz:] Das Hin– und Hernavigieren zwischen den Fenstern verläuft nicht erwartungsgemäß, Fenster lassen sich nicht schließen.
  `Lösung durch Navigation im Karteikasten.`

2.4.4. Farbgestaltung

- Farbe wird sparsam und nie als alleiniger Informationsträger eingesetzt.
  [Kreutz:] Ja.
  [Schulz:] In den Fenstern ‚Symptome des Falles' und ‚Wählen Sie bitte eine oder mehrere Informationen aus' ist ein Teil der notwendigen Information nur über die Farbe kodiert.
  `Die Kodierung über die Farbe wurden in diesem Fall gewählt, um die generelle Ordnung der Symptome in der Falldarstellung nicht zu stören.`

- Mit Rücksicht auf farbenblinde Benutzer werden Kombinationen wie rot / grün oder blau / violett, insbesondere für Texte und Symbole vermieden. Statt Mischfarben, die sich in Nuancen unterscheiden, werden „klare"Farben verwendet. Dabei empfiehlt sich eine Umrahmung der Farben mittels einer schwarzen Linie zur Verstärkung des Kontrasts.
  Ja und Farben sind auch individuell einstellbar.

- Farbensymbolik ist konsistent.
  [Kreutz:] Ja.
  [Leiner:] Rot wird sowohl für Symptome mit Symptomerkennung als auch für bereits durchgeführte Untersuchungen verwendet.
  [Schulz:] Ja.
  `Die beiden angesprochenen Rotkodierungen befinden sich in unterschiedlichen Fenstern (also auch inhaltliche anderen Bereichen), können aber durch Experten oder Lerner jederzeit individuell verändert werden.`

2.4.5. Hilfefunktion

- Bedienungslogik und Bedienelemente werden in einer Online–Hilfe erläutert.
  Keine Online–Hilfe

- Vorhandensein eines gedruckten Handbuchs ist kein Qualitätskriterium an sich. Wo eines existiert, ist es verständlich und ansprechend gestaltet, gut strukturiert mit Inhaltsverzeichnis und verfügt über ein Stichwortregister.
  Kein Inhaltsverzeichnis, kein Stichwortregister, aber ausführlich und verständlich geschrieben.

2.5. Dialog und Didaktik

- Das Lernpensum ist inhaltlich und bezüglich einer zeitlichen Abschätzung klar umrissen, der Lernstoff ist modular gegliedert.
  [Leiner:] Ja, durch die fallweise Bearbeitung.
  [Schulz:] Über die notwendige Zeit wird keine Aussage gemacht.

Da der D3Trainer hauptsächlich zum Einüben und Überprüfen von vorhandenem Wissen eingesetzt wird, können die nächsten Punkte des Kriterienkataloges nur partiell angewendet werden.

2.5.2. Überprüfung des vermittelten Wissens

- Lerndialoge beschränken sich nicht nur auf geschlossene Fragen (Multiple–Choice–Aufgaben, Objektmarkierungen), sondern umfassen auch offene Fragen (Freitext).
  Keine offenen Fragen.
  Da es sich hier meist um MC-Fragen mit ca. 200 Antwortalternativen handelt, kann von einer Zwischenstufe zwischen offenen und geschlossenen Fragen gesprochen werden.

- Die Bewertung von Benutzerantworten ist konstruktiv, d. h. es wird eine Erläuterung angeboten.
  [Kreutz:] Ja.
  [Leiner:] Ja, auch wenn die Erläuterung bei der Zielgruppe zu Verständnisproblemen führen kann.
  Hier ist die technisch orientierte generierte Erklärung aus dem Wissensmodell gemeint. Erste Ansätze zur Verbesserung dieser Erklärungen werden in der nächsten Version berücksichtigt werden.

- Lerndialoge orientieren sich an konkreten Prüfungssituationen und in der beruflichen Praxis.
  Ja.

- Simulationen stellen, falls erforderlich unterstützt durch grafische und fotografische Elemente, die reale Situation möglichst realistisch

dar bzw. fokussieren die zu beachtenden oder zu verstehenden Elemente bzw. Vorgänge.
Ja, multimediale Elemente auf Text und Bild begrenzt.

- Visualisierungen und realitätsnahe Darstellungen sind so angelegt, daß sie vom Lernenden mit bestehenden Erfahrungen assoziiert werden können.
  Ja.

- In den Lernvorgang eingebettete Problemlösungsaufgaben sind geeignet, das erworbene Wissen anzuwenden und somit zu verfestigen.
  Ja.

- Konkrete Beispiele und Fallbeispiele erleichtern das Erlernen neuer Konzepte.
  Die gesamte Fallbearbeitung hat Problemlösungscharakter, umgesetzt anhand von Beispielen.

- Die Technik der Wissensüberprüfung ist dem Gesamtcharakter des Systems angepaßt und geht deswegen ggf. über rein textuelle Möglichkeiten hinaus. Werden beispielsweise verstärkt interaktive Techniken wie Simulationen zum Wissenserwerb eingesetzt, werden derartige Techniken auch bei der Wissensüberprüfung genutzt.
  Ja.

- Die Wissensüberprüfung orientiert sich an realen Prüfungsmodalitäten und –sequenzen.
  Ja.

- Bei Wissensabfragen wird folgenden Aspekten Rechnung getragen:

  – Positive Verstärkung
    Ja.

  – Korrektive Rückmeldungen mit Vertiefungsmöglichkeit
    Ja.

  – „Intelligente“Fragenpräsentation auch in Abhängigkeit des bisherigen Antwortverhaltens
    [Leiner:] Nein.
    `Da kein Studentenmodell gehalten wird, kann in der Bewertung auch kein Bezug darauf genommen werden. Das Feedback ist aber sehr detailliert, so daß es durch die Hinzunahme der Historie nicht wesentlich verbessert werden könnte. Man muß hier auch beachten, daß bei diesem komplexen Stoffgebiet die`

`Wiederholrate eines Wissenselements sehr gering ist.`

  - Überspringen von Fragen
    [Leiner:] Ja.
  - Möglichkeit, die richtige Lösung abzurufen
    [Leiner:] Ja.
  - Textuelle und grafische Rückmeldung des Lernfortschritts
    [Leiner:] Partiell für eine einzelne Aufgabe; am Ende eher pauschale Gesamtbewertung.

- Nach Abarbeitung der einzelnen Lerneinheiten ist eine bewertende Rückmeldung der Sitzung abrufbar.
  Ja.

2.5.3. Dialog / Navigation

- Bearbeitete Bereiche sind als solche markiert.
  [Kreutz:] Nein.
  [Schulz:] Teilweise.

- Trainingssitzungen können unterbrochen und wiederaufgenommen, aber auch vorzeitig beendet werden. Ein Wiederaufsetzpunkt nach einer Unterbrechung ist leicht ansteuerbar.
  Nein.

- Der Grad zweckmäßiger Selbststeuerung ist von der Zielgruppe abhängig. Für Anfänger wird eine „guided tour“ mit geringem Selbststeuerungsbedarf angeboten.
  [Leiner:] zweckmäßiger Selbsteuerungsgrad.

- Es besteht die Möglichkeit, den Grad der Selbststeuerung vorzugeben und während der Benutzung die Kompetenz zur Selbststeuerung schrittweise auf– oder abzubauen.
  [Leiner:] Zwei Hauptmodi.
  `Eigentlich vier Hauptmodi zur Symptomanforderung. Grundsätzlich sind aber alle Einstellungen zur Selbststeuerung in Abschnitt 3 genau beschrieben.`

2.5.4. Motivation

- Aktivierende, z. B. spielerische Elemente verhindern das Aufkommen von Monotonie.
  [Kreutz:] Fälle sind abwechslungsreich.
  [Schulz:] wenige (symbolhafte Handbewegungen)

`Die persönliche Vorstellung des Patienten durch eine kleine Geschichte entspricht grundsätzlich der Anfangsmotivierung in der problemorientierten Lehre. Durch den hohen Grad an Eigeninitiative sollte aber generell die Gefahr der Langweile bei der Fallbearbeitung gering sein.`

- Wettkampfähnliche Elemente (High–Score–Tabellen etc.) erhöhen die Motivation.
  [Schulz:] Ja, über das Fenster Fallaufwand.

- Dramaturgische Elemente werden motivationsfördernd eingesetzt (Einbettung in Rahmenhandlung, Simulation, Rollenspiel, . . .). Die Wahl dieser Elemente erfolgt in Anlehnung an das Kommunikationsverhalten der Zielgruppe.
  [Kreutz:] Ja, Rahmenhandlung.

2.6. Extensibility

Das System bietet dem Benutzer die Möglichkeit, Teile der Inhalte zu modifizieren und eigene Informationseinheiten (Fälle) zu ergänzen.
[Kreutz:] Nein.
[Leiner:] Nein, evtl. mit der Begründung eigener Vorschläge.
`Grundsätzlich kann in dem D3Trainer ja durch die Eingabe neuer Fälle in der Konsultation eine solche Ergänzung eingegeben werden. In der jetzigen Version des RheumaTutors ist dies jedoch nicht vorgesehen.`

Schlußbemerkung

Der Einsatz des GMDS–Kriterienkatalogs mit Unterstützung der verschiedenen Gutachter, von denen zwei bei der Entwicklung des Katalogs mitgewirkt haben (Schulz, Kreutz), zeigt, daß der Katalog zwar wichtige Aspekte bei der Entwicklung von CBT–Systemen beleuchtet, aber nicht universell auf auf alle Arten von Systemen gleich gut anwendbar ist. Trotzdem kann dieser Katalog bei jeder Art von System entwicklungsbegleitend eingesetzt werden und vor vielen Designfehlern bewahren. Für eine befriedigende Evaluation intelligenter fallbasierter Systeme müßte dieser aber noch erweitert werden.

Literaturverzeichnis

[AUHUBER 1998] AUHUBER, THOMAS C. (1998). *Entwicklung und Evaluation eines computerunterstützten Lernsystems in der Medizin – MicroPat – ein interaktiver Atlas der Histopathologie mit adaptierbaren Tutor*. Peter Lang, Europäische Hochschulschriften.

[BALCI 1997] BALCI, OSMAN (1997). *Principles of simulation model validation, verification and testing*. Transactions of the Society for Computer Simulation, 14:3–12.

[BERBERICH and BAMBERGER 1998] BERBERICH, MITCHEL and S. BAMBERGER (1998). *Building web-based knowledge clusters*. In *Proceedings of the Colloquium Web-based Knowledge Server of the IEEE*.

[BERGSTEINER 1995] BERGSTEINER, TORE (1995). *Entwurf und Implementierung eines sokratischen Dialogs*. Diplomarbeit, Universität Würzburg.

[BLOCK 1996] BLOCK, KARSTEN (1996). *Evaluation von Lernsoftware*. Informatik–Seminar, Universität Oldenburg http://www-cg-hci.informatik.uni-oldenburg.de/˜pgse96/Seminar/KBlock.html.

[ELIOT and WOOLF 1996] ELIOT, CHRISTOPHER and B. P. WOOLF (1996). *Iterative development and validation of a simulation-based medical tutor*. In FRASSON, CLAUDE, G. GAULTHIER, and A. LESGOLD, eds.: *Proceedings of the Third International Conference ITS'96*, vol. 1086 of *Lecture Notes in Computer Science*, pp. 540–549. Springer.

[ERNST 1996] ERNST, ROMAN (1996). *Untersuchung verschiedener Problemlösungsmethoden in einem Experten– und Tutorsystem zur makroskopischen Bestimmung krautiger Blütenpflanzen*. Diplomarbeit, Universität Würzburg, Biologie.

[FÄTH 1995] FÄTH, MARCO (1995). *Entwurf und Implementierung einer Simulationsshell zur tutoriellen Anwendung*. Diplomarbeit, Universität Würzburg.

[FAULHABER 1997] FAULHABER, SVEN (1997). *Entwicklung einer WWW-Oberfläche für den D3-Trainer*. Diplomarbeit, Universität Würzburg.

[FAULHABER und REINHARDT 1997] FAULHABER, SVEN und B. REINHARDT (1997). *D3-WWW-TRAINER Entwicklung einer Oberfläche für die Netzanwendung*. In: HERZOG, Hrsg.: *Workshop der GI: Intelligente Lehr/Lernsysteme*, Bd. tum-i9736, S. 31–40. Interner Bericht TU München.

[GAPPA 1995] GAPPA, UTE (1995). *Graphische Wissensakquisitionsysteme*. Doktorarbeit, Universität Karlsruhe. DISKI 100.

[GOOS 1995] GOOS, KLAUS (1995). *Fallbasiertes Schliessen*. Doktorarbeit, Universität Würzburg. DISKI.

[GOOS and SCHEWE 1993] GOOS, KLAUS and S. SCHEWE (1993). *Case based reasoning in clinic evaluation*. In *4th European Conference on Artificial Intelligence in Medicine AIME93*.

[IGLEZAKIS 1999] IGLEZAKIS, IONNIS (1999). *Kombination verschiedener Problemloeser*. In: PUPPE, FRANK, Hrsg.: *Workshop Proceedings zur XPS-99*. Springer.

[KLÜGL and PUPPE 1998] KLÜGL, FRANZISKA and F. PUPPE (1998). *The multi–agent simulation environment sesam*. In BÜNING, H. KLEINE, ed.: *Proceedings des Workshops Simulation in Knowledge–based Systems*.

[KOHLERT 1995] KOHLERT, SIGFRIED (1995). *Entwurf und Implementierung eines minimalen funktionalen Problemlösers*. Diplomarbeit, Universität Würzburg.

[LANGER 1997] LANGER, IRIS (1997). *Lernen am Computer: Evaluation eines intelligenten Tutorsystems im universitären Einsatz*. Diplomarbeit, Universität Würzburg.

[LEINER et al. 1999] LEINER, FLORIAN, S. SCHEWE, C. BETZ, J. PRINZ, F. PUPPE, B. REINHARDT und O. RIENHOFF (1999). *Ein Modellvorhaben zur computergestützten ärztlichen Fortbildung mit Steuerungs- und Evaluationsmechanismen*. In: *44. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS)*.

[MARK and GREER 1993] MARK, MARY A. and J. E. GREER (1993). *Evaluation methodologies for intelligent tutoring systems*. Journal of Artificial Intelligence in Education, 4(3).

[POECK 1995] POECK, KARSTEN (1995). *Konfigurierbare Problemlösungsmethoden am Beispiel der Problemklassen Zuordnung und Diagnostik*. Doktorarbeit, Universität Würzburg. DISKI 86.

[POECK 1994] POECK, KLAUS (1994). *Neurologie*. Springer, 9. Aufl.

[POECK und HACKE 1998] POECK, KLAUS und W. HACKE (1998). *Neurologie*. Springer, 10. Aufl.

[Psycho 1990] PSYCHO (1990). *Wörterbuch der Psychologie*. dtv.

[PUPPE 1998] PUPPE, FRANK (1998). *Meta knowledge for extending diagnostic consultation to critiquing systems*. In *Proceedings of EKAW–99 11th workshop on Knowledge Acquisition, Modeling, and Management*. Springer.

[PUPPE et al. 1996] PUPPE, FRANK, U. GAPPA, K. POECK und S. BAMBERGER (1996). *Wissensbasierte Diagnose- und Informationssysteme*. Springer.

[PUPPE et al. 1998] PUPPE, FRANK, B. PUPPE, B. REINHARDT, S. SCHEWE und H.-P. BUSCHER (1998). *Evaluation medizinischer Diagnostik-Expertensysteme zur Wissensvermittlung*. Informatik Biometrie und Epidemiologie in Medizin und Biologie, 29(1):48–59.

[PUPPE and REINHARDT 1995] PUPPE, FRANK and B. REINHARDT (1995). *Generating case-oriented training from diagnostic expert systems*. Machine Mediated Learning, 5(4):199–219.

[PUPPE et al. 1995] PUPPE, FRANK, B. REINHARDT, and K. POECK (1995). *Generated critic in the knowledge–based neurology trainer*. In *Proceedings of the Conference on Artificial Intelligence in Medicine in Europe (AIME 95)*, pp. 427–428.

[REINHARDT 1994] REINHARDT, BETTINA (1994). *Symptomerkennung in einem multimedialen Trainingssystem*. In: SCHOOF, ERIC und GLOWALLA, Hrsg.: *Hypermedia in Aus- und Weiterbildung*.

[REINHARDT 1996] REINHARDT, BETTINA (1996). *Expert systems and hypertext for teaching diagnostics*. In BRNA, PAUL, A. PAIVA, and J. SELF, eds.: *Proceedings of the European Conference on Artificial Intelligence in Education*, pp. 298–304. Edicoes Colibri.

[REINHARDT 1997] REINHARDT, BETTINA (1997). *Generating case oriented intelligent tutoring systems*. In *Technical Report*, vol. FS-97-01 of *Papers from the 1997 AAAI Fall Symposium*, pp. 79–85. AAAI Press.

[REINHARDT und PUPPE 1997] REINHARDT, BETTINA und F. PUPPE (1997). *Didaktische Aspekte in fallorientierten intelligenten Trainingssystemen*. In: CONRADI, HELMAR, R. KREUTZ und K. SPITZER, Hrsg.: *CBT in der Medizin - Methoden, Techniken, Anwendungen: Proc. GMDS Workshop in Aachen on Computer Based Training*, S. 157–168. GMDS, Verlag der Augustinus-Buchhandlung.

[REINHARDT und REMP 1995] REINHARDT, BETTINA und T. REMP (1995). *Integration von Hypertext und Expertensystemen am Beispiel eines Trainingssystems für Herz–Rhythmus–Störungen*. In: HUBER-WÄSCHLE, FRIEDBERT, H. SCHAUER und P. WIDMAYER, Hrsg.: *GISI 95 - Herausforderung eines globalen Informationssverbundes für die Informatik*, S. 275–283. Springer Informatik aktuell.

[REINHARDT and SCHEWE 1995] REINHARDT, BETTINA and S. SCHEWE (1995). *A shell for intelligent tutoring systems*. In GREER, JIM, ed.: *Proceedings of the World Conference on Artificial Intelligence in Education*, pp. 83–90. Association for the Advancement of Computing in Education, Association for the Advancement of Computing in Education.

[RIENHOFF 1998] RIENHOFF, O. (1998). *Ärztliche Fortbildung und Qualitätssicherung*. In: *Forum Info 2000*, Bd. 105 d. Reihe *Schriftenreihe des Bundesministeriums für Gesundheit*, S. 114–117. Bundesministeriums für Gesundheit.

[SCHEWE et al. 1996] SCHEWE, STEFAN, T. QUAK, B. REINHARDT, and F. PUPPE (1996). *Evaluation of a knowledge-based tutorial program in rheumatology*. In FRASSON, CLAUDE, G. GAULTHIER, and A. LESGOLD, eds.: *Proceedings of the Third International Conference ITS'96*, vol. 1086 of *Lecture Notes in Computer Science*, pp. 531–539. Springer.

[SCHEWE et al. 1999] SCHEWE, STEFAN, B. REINHARDT, and C. BETZ (1999). *Experiences with a knowledge–based tutoring system for student education in rheumatology*. In *XPS–99: Wissensbasierte Systeme – Bilanz und Perspektiven*, Lecture Notes of Artificial Intelligence. Springer.

[SCHEWE and SCHREIBER 1993] SCHEWE, STEFAN and M. SCHREIBER (1993). *Stepwise development of clinical expert system in rheumatology*. The Clinical Investigator, 71:139–144.

[SCHULZ et al. 1997] SCHULZ, STEFAN, T. AUHUBER und R. KLAR (1997). *Kontrollierte Evaluationsstudie von MicroPat – Interaktiver Atlas der Histopathologie*. In: CONRADI, HELMAR, R. KREUTZ und K. SPITZER, Hrsg.: *CBT in der Medizin - Methoden, Techniken, Anwendungen*, S. 117–123. GMDS, Verlag der Augustinus-Buchhandlung.

[SEITZ et al. 1999] SEITZ, ALEXANDER, A. MARTENS, J. BERNAUER, C. SCHEUERER, and J. THOMSEN (1999). *An architecture for intelligent support of authoring and tutoring in medical multimedia learning environments*. In *Proceedings to ED–Media 1999*.

[SERVE 1992] SERVE, HELMUT J. (1992). *Das Unterrichtsprinzip der Motivierung*. In: SEIBERT, NORBERT und H. J. SERVE, Hrsg.: *Prinzipien guten Unterrichts*. PIMS Verlag, München.

[SETTELE 1998] SETTELE, CLAUDIA (1998). *Autorensystem zur Präsentation und Abfrage von informellem Wissen*. Diplomarbeit, Universität Würzburg.

[SHUTE and REGIAN 1993] SHUTE, VALERIE J. and J. W. REGIAN (1993). *Principles for evaluating intelligent tutoring systems*. Journal of Artificial Intelligence in Education, 4(3):245–271.

[TOLKSDORF 1998] TOLKSDORF, MARKUS (1998). *Entwurf und Implementierung einer Shell für Bayessche Netzwerke*. Diplomarbeit, Universität Würzburg.

[WENGER 1987] WENGER, ETIENNE (1987). *Artificial intelligence and tutoring systems*. Morgan Kaufman.

Veröffentlichungen aus dem Projekt

Betz, C. *Information Generation and Navigation in Problem Based Training Systems*, in Proc. Workshop Adaptive and Intelligent Web-Based Education Systems, ITS'2000 - The New Conditions of Learning and Training, Montreal, 2000.

Betz, C. *Informationsrecherche in problembasierten Tutorsystemen*, in Proc. 5. Workshop der AG CBT in der Medizin der GMDS, Köln, 2000.

Betz, C. *Problemorientierte Trainingssysteme im World Wide Web*, in Proc. 10. Arbeitstreffen der GI-Fachgruppe 1.1.5/7.0.1, Hamburg, 2000.

Faulhaber S., Reinhardt B. *D3-WWW-TRAINER Entwicklung einer Oberfläche für die Netzanwendung*, Workshop der GI: Intelligente Lehr/Lernsysteme September 97 in Duisburg, Interner Bericht der TU München TUM-I9736,31-40, August 1997.

Puppe, F., Reinhardt, B., Schultheis, K. *Computerunterstütztes problemorientiertes Lernen*, in Desel, J. (Hrsg.): Was ist Informatik?, Springer, 135-156, 2000.

Puppe, F. *Evidence-based Medicine for Defining Medical Knowledge Bases*, in Bruch, H.-P., Köckerling, F., Bouchard, R., Schug-Paß, C: New Aspects fo High Technology in Medicine, Monduzzi Editore, 435-442, 2000.

Puppe, F. und Reinhardt, B. *Diagnostische Modelle und Tutorsysteme zur Unterstützung ärztlicher Entscheidungsfindung*, in: Fischer, M. und Bartens, W. (Hrsg.): Zwischen Erfahrung und Beweis: Medizinische Entscheidungen und Evidence-bases Medicine, Verlag Hans Huber, 175-206, 1999.

Puppe F., Puppe B., Reinhardt B. Schewe S., Buscher H.P. *Evaluation medizinischer Diagnostik-Expertensysteme zur Wissensvermittlung.*, Informatik, Biometrie und Epidemiologie in Medizin und Biologie 29 (1), 48-59, 1998.

Puppe, F. und Schewe, S. *Mehrfachverwendung von diagnostischen Wissensbasen in der Medizin*, Künstliche Intelligenz 97 / 3, 15-23, 1997.

Reinhardt, B. *Didaktische Strategien in generierten Trainingssystemen zum diagnostischen Problemlösen*, Dissertation, erschienen im infix-Verlag, Diski 234, 2000.

Reinhardt B. *Kritikoptionen in fallbasierten Trainingssystemen*, in Proc. GMDS-Workshop in Heidelberg on Computer Based Training, 1999.

Reinhardt, B. und Puppe, F. *Generierung flexibler intelligenter Trainingssysteme*, in: Künstliche Intelligenz Heft 4 / 1999, 12-18, 1999.

Reinhardt B. *Intelligente Autorensysteme für Computer-Basiertes Training*, Workshop Intelligente Unterstützung für offene Lehr / Lernsysteme, Olaf Schröder (Hrsg.), September 98 in Bremen, 22. Jahrestagung der Künstlichen Intelligenz (KI-98), TZI-Bericht Nr. 8, 1998.

Reinhardt B. *Generating Case Oriented Intelligent Tutoring Systems*, AAAI Fall Symposium ITS Authoring Systems November 97 in Boston, 1997.

Reinhardt B., Puppe F. *Didaktische Aspekte in fallorientierten intelligenten Trainingssystemen*, in Conradi, H., Kreutz, R. und Spitzer, K. (Hrsg.): CBT in der Medizin - Methoden Techniken, Anwendungen, Proc. GMDS-Workshop in Aachen on Computer Based Training, Verlag der Augustinus-Buchhandlung, 157-168, 1997.

Schewe S., Reinhardt B., Betz C. *Experiences with a Knowledge Based Tutoring System for Student Education in Rheumatology*, in Puppe, F. (ed.): XPS-99: 5th German Conference on Knowledge-Based Systems, Springer, LNAI 1570, pp.193-200, 1999.